

How Clustering Affects Epidemics in Random Networks

Emilie Coupechoux, *INRIA - ENS*

Marc Lelarge, *INRIA - ENS*

E-mail: Emilie.Coupechoux, Marc.Lelarge @ens.fr

February 23, 2012

Abstract

Motivated by the analysis of social networks, we study a model of random networks that has both a given degree distribution and a tunable clustering coefficient. We consider two types of growth processes on these graphs: diffusion and symmetric threshold model. The diffusion process is inspired from epidemic models. It is characterized by an infection probability, each neighbor transmitting the epidemic independently. In the symmetric threshold process, the interactions are still local but the propagation rule is governed by a threshold (that might vary among the different nodes). An interesting example of symmetric threshold process is the contagion process, which is inspired by a simple coordination game played on the network. Both types of processes have been used to model spread of new ideas, technologies, viruses or worms and results have been obtained for random graphs with no clustering. In this paper, we are able to analyze the impact of clustering on the growth processes. While clustering inhibits the diffusion process, its impact for the contagion process is more subtle and depends on the connectivity of the graph: in a low connectivity regime, clustering also inhibits the contagion, while in a high connectivity regime, clustering favors the appearance of global cascades but reduces their size.

For both diffusion and symmetric threshold models, we characterize conditions under which global cascades are possible and compute their size explicitly, as a function of the degree distribution and the clustering coefficient. Our results are applied to regular or power-law graphs with exponential cutoff and shed new light on the impact of clustering.

Keywords: Contagion threshold, Diffusion, Random graphs, Clustering

1 Introduction

Many network phenomena are well modeled as spreads of epidemics through a network. However, depending on the motivation, different mechanisms are at work and we believe that different models should be used. For example, to model the spread of worms or email viruses, and, more generally, faults, a simple diffusion model is often used, where each node when it becomes infected or faulty 'transmits' the infection to her neighbors independently and with a given probability. There is now a vast literature on such epidemics on complex networks (see [24] for a review). But if nodes represent agents in a social network, the transmission mechanism is independent of the local condition faced by the agents concerned. But to model the spread of innovations as a social process, individual's adoption behavior is highly correlated with the behavior of her contacts. In this case, there is a factor of persuasion or coordination involved and relative considerations tend to be important in understanding whether some new behavior or belief is adopted [27]. In social contexts, the spread of information and behavior often exhibits features that do not match well those of the diffusion model just described where an individual is influenced by each of her neighbors independently. In such a context, the linear threshold model, originally proposed by

Granovetter [12], captures in a simple way the local correlations among individuals. We will study the symmetric threshold model proposed by Lelarge [20], which generalizes both bootstrap percolation [3] and contagion model [22].

In this paper, we will analyze two different types of epidemics modeled by simple growth processes that we now describe. In a given network, each node can either be active or inactive. The *diffusion* process corresponds to the case where each node of the network that becomes active transmits the activation to her neighbors with a given probability, independently from each others. On the other hand, in the *symmetric threshold* model, a threshold is associated to each node, and the dynamics of the process corresponds to the case where a node of the network becomes active as soon as the number of her active neighbors exceeds the threshold of the node. As in the original model of [20], thresholds are (possibly) random, with a distribution depending on the degree of the node, and such that thresholds are independent among nodes. The symmetric threshold model will allow us to analyze the contagion process [22]. For both models, we will first consider a case where there is only one initial active node and characterize conditions under which global cascades are possible, that is to say when a positive fraction of the population is active at the end of the process. In such cases we compute the probability of a cascade and its size. Then, we consider the cascade size when a positive fraction of the population is initially active. The initial activations are random in that case, and the probability that a node belongs to the seed might depend on its degree.

We now describe the model of random graphs studied in this paper. For many real-world networks, the underlying graph G is a power-law graph, *i.e.* a graph whose degree distribution follows a power law. Random graphs with a given degree sequence allow to model such behavior. This model is usually called the configuration model [5]. There is a vast literature on the analysis of the diffusion for such graphs [24]. The contagion process has also been studied for such graphs through heuristics [28] or rigorously in [20] or [2]. Random graphs are not considered to be highly realistic models of most real-world networks, and they are used as first approximation as they are a natural choice for sparse interaction network in the absence of any known geometry. One essential drawback of this model is that these graphs are 'locally tree-like': short cycles are very rare. However, real-world networks are often highly clustered, meaning that there is a large number of triangles and other short cycles [23]. For social graphs, this is a consequence of the fact that friendship circles are typically strongly overlapping so that many of our friends are also friends of each other.

There are works in the physics and biology literature on models of random graphs with clustering [25]. Our model is inspired from [26] which allow to model random graphs with positive clustering and possibly power law degree distribution. The idea is to 'add' clustering to a standard configuration model by replacing some vertices by cliques. By choosing the fraction of vertices replaced, this leads to a graph where the amount of clustering can be tuned by adjusting the parameters of the model. This model generalizes the standard configuration model to incorporate clustering. Understanding how clustering affects diffusion and contagion remains largely an open question. Our work is a first step towards addressing this issue in a systematic and rigorous way. In particular, we are able to make a rigorous analysis of the impact of a variation of the clustering coefficient while keeping the degree distribution in the graph fixed. To the best of our knowledge, this sensitivity analysis is new and gives a number of insights on the impact of clustering. In [26] and [11], the diffusion process on such graphs is analyzed by an heuristic approximation through a branching process with additional cliques. We derive rigorous proofs for these results. A different model of random graphs with clustering, called random intersection graphs has been studied rigorously in [9] and in [7], the diffusion process is studied on such graphs. However, the degree distribution for this kind of graphs has to be a Poisson distribution and clustering cannot vary independently of the degree distribution. Up to our knowledge, results on the contagion model have not been proved before our work for random graphs with clustering.¹ Recently, [1] derives bounds which are valid for the contagion model [22] on deterministic networks. Our analysis in contrast gives asymptotic results as the size of the graph tends to infinity and allows us (by looking at a more specific model) to identify neatly the impact of clustering.

¹ A preliminary version without proofs of our work appeared in [8].

The paper is organized as follows. In Section 2, we present the graph model, compute its asymptotic degree distribution and its asymptotic clustering coefficient. We explain how to tune the clustering coefficient of the model, while keeping the asymptotic degree distribution fixed. In Section 3, we derive the minimal value for the probability of infection in our random graph model with clustering such that a global diffusion is possible, and we compute the size of the diffusion in that case. We apply these results to random regular graphs and power-law graphs and show that clustering inhibits the diffusion process in that case. We also compute the cascade size for the diffusion process with degree based activation. In Section 4, we derive the cascade condition for the symmetric threshold model on our random graph model with clustering, together with the size of the cascade when it occurs. Numerical evaluations in the particular case of the contagion process show that the effect of clustering on the contagion threshold depends on the mean degree of the graph, and that clustering decreases the cascade size when it occurs. In addition, we finally compute the cascade size for the symmetric threshold model with a slight variant of the degree based activation. Proofs are given in Section 5.

Notations. In the following, we consider asymptotics as $n \rightarrow \infty$, and we denote by \rightarrow_p the convergence in probability as $n \rightarrow \infty$. The abbreviation ‘whp’ (‘‘with high probability’’) means with probability tending to 1 as $n \rightarrow \infty$, and we use the notation $o_p(n)$, $\Theta_p(n)$ in a standard way (see [15] for instance): $X = o_p(n)$ means that, for every $\varepsilon > 0$, $\mathbb{P}(X > \varepsilon n) \rightarrow 0$ as $n \rightarrow \infty$. In addition, for integers $s \geq 0$ and $0 \leq r \leq s$, let b_{sr} denote the binomial probabilities $b_{sr}(p) := \mathbb{P}(\text{Bi}(s, p) = r) = \binom{s}{r} p^r (1-p)^{s-r}$.

2 Random graph model and its basic properties

We first present the model for the random graph, and compute its asymptotic degree distribution and its asymptotic clustering coefficient (for two different definitions).

2.1 Model of random graph with clustering

We first consider the uniform random graph with fixed degree distribution: since this graph has asymptotically no clustering, we will then modify it to obtain a graph with clustering.

Let $n \in \mathbb{N}$ and let $\mathbf{d} = (d_i^{(n)})_{i=1}^n = (d_i)_1^n$ be a sequence of non-negative integers such that $\sum_i d_i$ is even. The integer n is the number of vertices in the graph and vertex $i \in [n]$ has degree d_i in the graph. Let $G(n, \mathbf{d})$ be a graph chosen uniformly at random among all simple (i.e. with no multi-edges or self-loops) graphs with n vertices and degree sequence \mathbf{d} (assuming such graphs exist) [5].

We will let $n \rightarrow \infty$ and assume that we are given \mathbf{d} satisfying the following regularity conditions which are standard in the random graph literature, see [21]:

Condition 1. For each n , $\mathbf{d} = (d_i)_1^n$ is a sequence of non-negative integers such that $\sum_i d_i$ is even. We assume that there exists a probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$ (independent of n) such that:

- (i) $n_r/n = |\{i : d_i = r\}|/n \rightarrow p_r$ as $n \rightarrow \infty$, for all $r \geq 0$;
- (ii) $\lambda := \sum_r r p_r \in (0, \infty)$;
- (iii) $\sum_i d_i^3 = O(n)$.

If D_n is the degree of a vertex chosen uniformly at random among the n vertices of $G(n, \mathbf{d})$, and D a random variable with distribution \mathbf{p} , (i) is equivalent to the fact that $D_n \xrightarrow{d} D$ (convergence in distribution). In addition, (iii) is equivalent to $\mathbb{E}[D_n^3] = O(1)$, which implies that the random variables D_n are uniformly integrable or equivalently the uniform summability of $\sum_d d n_d/n$, in particular $\mathbb{E}[D_n] \rightarrow \mathbb{E}[D]$.

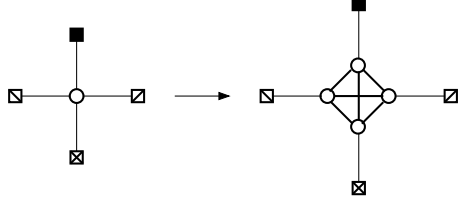


Figure 1: Transformation

The model of random graphs $G(n, \mathbf{d})$ are 'locally tree-like', i.e. they contain very few (i.e. $o(n)$) short cycles in their structure. We now show that it is possible to generalize this model of random graphs to incorporate clustering in a simple way. The resulting model of random graphs will still be tractable for the analysis of diffusion and symmetric threshold models. To 'add' clustering in $G(n, \mathbf{d})$, we replace some vertices by a clique of size the degree in the original graph, i.e. a vertex of degree r in the original graph $G(n, \mathbf{d})$ is replaced by r vertices with all the $r(r-1)/2$ edges between them and each of them is connected to exactly one of the neighbors of the vertex in the original graph $G(n, \mathbf{d})$ as illustrated on Figure 1. Note that if $r = 0$, i.e. if the original node is isolated, this procedure remove the node. By convention a clique of size zero is empty.

In order to be able to tune the clustering coefficient in the graph, we will not replace all vertices by a clique but do a probabilistic choice whether to replace a vertex or not: for all $r \geq 0$, $\gamma_r \in [0, 1]$ represents the probability that a vertex of degree r in $G(n, \mathbf{d})$ is replaced by a clique of size r in the new model denoted $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = (\gamma_r)_{r=0}^\infty$ is a short notation for the sequence of γ_r 's. The choices to replace or not a vertex by a clique are made independently at each vertex of the original graph $G(n, \mathbf{d})$. More formally, for each vertex $i \in \{1, \dots, n\}$, let $X(i)$ be a Bernoulli random variable with parameter γ_{d_i} (all Bernoulli random variables being independent of each other). We construct the random graph $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$ by replacing each vertex i of $G(n, \mathbf{d})$ with $X(i) = 1$, by a clique of size d_i where each vertex of the clique has exactly one neighbor outside the clique being a neighbor of i in the original graph. All vertices i with $X(i) = 0$ are unchanged. In particular, if $\gamma_r = 0$ for all $r \geq 0$, then we simply get $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma}) = G(n, \mathbf{d})$, whereas if $\gamma_r = 1$ for all $r \geq 0$, all vertices in $G(n, \mathbf{d})$ are replaced by cliques (and isolated vertices are removed). With a little abuse of notation, we write $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$ for the graph $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$ in which the sequence $\boldsymbol{\gamma}$ is constant and equals to $\boldsymbol{\gamma}$.

2.2 Degree distribution in $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$

As we will see in the next subsection, the procedure described above introduces clustering as soon as $\gamma_r > 0$ for some r . It also modifies the degree distribution in the graph and we derive the new degree distribution here. Recall that each vertex of degree r can either be replaced (with probability γ_r) by r vertices of degree r , or stays as a single vertex of degree r (with probability $1 - \gamma_r$). The following proposition gives the resulting asymptotic degree distribution in $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$.

Proposition 2. *We consider the model $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$ for a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$, and clustering parameter $\boldsymbol{\gamma} = (\gamma_r)_{r=0}^\infty$. For all $r \geq 0$, let \tilde{n}_r be the number of vertices with degree r in $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$, and let $\tilde{n} = \sum_r \tilde{n}_r$ be the total number of vertices in $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$. Then we have, as $n \rightarrow \infty$:*

$$\frac{\tilde{n}}{n} \xrightarrow{p} \tilde{\gamma} := \sum_{d \geq 0} [d\gamma_d + (1 - \gamma_d)] p_d > 0.$$

and, for all $r \geq 0$, the proportion of vertices with degree r in $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$ has the following limit, as $n \rightarrow \infty$:

$$\frac{\tilde{n}_r}{\tilde{n}} \xrightarrow{p} \tilde{p}_r := \frac{[r\gamma_r + (1 - \gamma_r)] p_r}{\tilde{\gamma}}.$$

Proof. We first show the asymptotics for the number \tilde{n} of vertices in $\tilde{G}(n, \mathbf{d}, \gamma)$.

Let $d \geq 0$, and let B_d be the number of vertices with degree d that are replaced by a clique. Then B_d follows a Binomial distribution with parameters (n_d, γ_d) . By Condition 1-(i), we have: $n_d/n \rightarrow p_d$, so that the Law of Large Numbers implies: $B_d/n \rightarrow_p \gamma_d p_d$ (which is still true if $p_d = 0$).

The number of vertices with degree d that are *not* replaced by a clique is $n_d - B_d$, so we can express the total number \tilde{n} of vertices in $\tilde{G}(n, \mathbf{d}, \gamma)$ the following way:

$$\frac{\tilde{n}}{n} = \frac{1}{n} \sum_d dB_d + (n_d - B_d) \xrightarrow{p} \sum_d [d\gamma_d + (1 - \gamma_d)] p_d = \tilde{\gamma}$$

which follows from the previous limits, and the uniform summability of $\sum dn_d/n$ implied by Condition 1-(iii).

Similarly, the total number of vertices with degree r ($r \geq 0$) in $\tilde{G}(n, \mathbf{d}, \gamma)$ is $rB_r + (n_r - B_r)$, so the proportion of vertices with degree r in $\tilde{G}(n, \mathbf{d}, \gamma)$ is:

$$\frac{rB_r + (n_r - B_r)}{\tilde{n}} \xrightarrow{p} \frac{[r\gamma_r + (1 - \gamma_r)] p_r}{\tilde{\gamma}} = \tilde{p}_r,$$

which concludes the proof. \square

In other words, if \tilde{D}_n is the degree of a vertex chosen uniformly at random in $\tilde{G}(n, \mathbf{d}, \gamma)$, then Proposition 2 implies that $\tilde{D}_n \xrightarrow{d} \tilde{D}$, where \tilde{D} is a random variable with distribution $(\tilde{p}_r)_{r \geq 0}$. In the particular case where $\gamma_r = \gamma$ for all r , we have $\tilde{\gamma} = \gamma\lambda + 1 - \gamma$ and the mean degree of the graph is then

$$\tilde{\lambda} = \mathbb{E}[\tilde{D}] = \frac{\gamma \mathbb{E}[D(D-1)] + \mathbb{E}[D]}{\gamma \mathbb{E}[D-1] + 1},$$

which is a non-decreasing function of γ .

2.3 Clustering coefficient

We now compute the clustering coefficient of the graph $\tilde{G}(n, \mathbf{d}, \gamma)$. The most common definition of the clustering coefficient of a finite graph is given by:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}} \in [0, 1]. \quad (1)$$

In our model of random graphs where vertices are exchangeable, this definition can also be interpreted as the conditional probability that there is an edge between two vertices j and k , given that they have a common neighbor i . At the end of this subsection, we will also consider a local clustering coefficient defined for a vertex of degree larger than 3 and compute the associated clustering coefficient for the graph based on this local measure for the graph $\tilde{G}(n, \mathbf{d}, \gamma)$.

Computation of the clustering coefficient. Note that the number of connected triples in $\tilde{G}(n, \mathbf{d}, \gamma)$ is simply $\sum_v d_v(d_v - 1)/2$. On the other hand, for any vertex v in $\tilde{G}(n, \mathbf{d}, \gamma)$, let P_v be the number of pairs of neighbors of v that share an edge together. More precisely, if \mathcal{N}_v is the set of neighbors of v (whose cardinality is $|\mathcal{N}_v| = d_v$), then P_v is the number of pairs $\{w, w'\} \subset \mathcal{N}_v$, such that w and w' are also neighbors of each other. Thus 3 times the number of triangles in $\tilde{G}(n, \mathbf{d}, \gamma)$ is given by $\sum_v P_v$. Hence following (1), we define the clustering coefficient $C^{(n)}$ of the graph $\tilde{G}(n, \mathbf{d}, \gamma)$ by:

$$C^{(n)} = \frac{2 \cdot \sum_v P_v}{\sum_v d_v(d_v - 1)} \in [0, 1].$$

Proposition 3. *We consider the model $\tilde{G}(n, \mathbf{d}, \gamma)$ for a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$, and clustering parameter $\gamma = (\gamma_r)_{r=0}^\infty$. Then we have for the clustering coefficient of $\tilde{G}(n, \mathbf{d}, \gamma)$:*

$$C^{(n)} \xrightarrow{p} C := \frac{\sum_{r \geq 2} r(r-1)(r-2)\gamma_r p_r}{\sum_{r \geq 2} ((r-1)\gamma_r + 1)r(r-1)p_r}.$$

The proof is given at the end of the subsection. We will explore in more details the implications of Proposition 3 in Section 2.4. Before that we present an alternative definition of clustering.

Another definition of clustering coefficient. The local clustering coefficient $C_v^{(n)}$ of a vertex v in a graph quantifies how close the vertex and its neighbors are to being a clique. $C_v^{(n)}$ is defined to be the fraction of pairs of neighbors of v that are also neighbors of each other [29]. Using the notations of the previous paragraph, the local clustering coefficient of v is $C_v^{(n)} = P_v \cdot 2/[d_v(d_v - 1)]$. Note that this definition only makes sense if $d_v \geq 2$, i.e. when v is not isolated, nor a leaf of the graph. The local clustering of vertices with degree one or zero is taken to be zero. The clustering coefficient for the whole network $C_2^{(n)}$ is then defined as the average of the clustering coefficients for each vertex: $C_2^{(n)} = \sum_v C_v^{(n)} / \tilde{n}$, where \tilde{n} is the number of vertices in the graph, including those of degree one or zero. As observed in [16], since we take the convention that the local clustering is zero for vertices with degree one or zero, the clustering coefficient in the graph can be very low if the graph contains a lot of such vertices, even if other vertices are highly clustered. We call $C_2^{(n)}$ the biased clustering coefficient and we refer to [16] for more information. In the next proposition, we give the asymptotics for the biased clustering coefficient in $\tilde{G}(n, \mathbf{d}, \gamma)$ but we will mainly deal with the definition (1) in the rest of the paper.

Proposition 4. *We consider the model $\tilde{G}(n, \mathbf{d}, \gamma)$ for a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$. Then we have for the biased clustering coefficient of $\tilde{G}(n, \mathbf{d}, \gamma)$:*

$$C_2^{(n)} \xrightarrow{p} C_2 := \sum_{r \geq 3} p_r \frac{\gamma_r}{\tilde{\gamma}} (r-2),$$

where $\tilde{\gamma}$ is defined in Proposition 2.

Proof of Propositions 3 and 4. We recall the following standard result for the random graph $G(n, \mathbf{d})$:

Lemma 5. *Let $\bar{C}^{(n)}$ (resp. $\bar{C}_2^{(n)}$) be the (resp. biased) clustering coefficient in $G(n, \mathbf{d})$. Then we have: $\bar{C}^{(n)} \xrightarrow{p} 0$ and $\bar{C}_2^{(n)} \xrightarrow{p} 0$.*

We say that a vertex in $\tilde{G}(n, \mathbf{d}, \gamma)$ has *parent* $i \in \{1, \dots, n\}$ if it belongs to a clique that replaces the vertex i of $G(n, \mathbf{d})$ (when $X(i) = 1$) or if it is i (when $X(i) = 0$). We first consider a vertex v in $\tilde{G}(n, \mathbf{d}, \gamma)$ whose parent i is such that $X(i) = 1$. In this case we can directly compute the local clustering coefficient $C_v^{(n)}$. Indeed, vertex v has $d_i - 1$ neighbors inside K , that are all linked together (which gives $\frac{(d_i-1)(d_i-2)}{2}$ edges in total), and one neighbor v' outside K , which is not linked to the other neighbors of v (if it were the case, there would be multiple edges between i and the parent j of v' , which is not the case in the simple graph $G(n, \mathbf{d})$). Hence

$$P_v = \frac{(d_i-1)(d_i-2)}{2} \quad \text{and} \quad C_v^{(n)} = \frac{2P_v}{d_i(d_i-1)} = \frac{d_i-2}{d_i},$$

provided that $d_i \geq 2$. If $d_i \in \{0, 1\}$, then $C_v^{(n)} = 0$.

We first prove Proposition 4. Since there are d_i such vertices inside a clique, the contribution of clique K in the total clustering $C_2^{(n)} = \sum_v C_v^{(n)} / \tilde{n}$ is equal to $d_i C_v^{(n)} / \tilde{n} = (d_i - 2) / \tilde{n}$. This leads to the following:

$$\frac{\tilde{n}}{n} C_2^{(n)} = \frac{1}{n} \sum_{d \geq 2} (d-2) B_d + \frac{1}{n} \sum_{i: X(i)=0} C_i^{(n)}$$

where B_d is the number of vertices with degree d that are replaced by a clique, as in the proof of Proposition 2. Using that $B_d/n \rightarrow_p \gamma_d p_d$, and that $\sum_{i: X(i)=0} C_i^{(n)}/n \rightarrow_p 0$ (as a consequence of Lemma 5), we obtain: $\frac{\tilde{n}}{n} C_2^{(n)} \xrightarrow{p} \sum_{d \geq 3} (d-2) \gamma_d p_d$. Proposition 4 follows, applying Proposition 2.

The end of the proof for Proposition 3 is similar, and follows from the fact that:

$$\begin{aligned} 2 \sum_v P_v/n &\rightarrow_p \sum_d d(d-1)(d-2) \gamma_d p_d, \\ \sum_v d_v(d_v-1)/n &\rightarrow_p \sum_d d(d-1)[d\gamma_d + (1-\gamma_d)]p_d. \end{aligned}$$

□

2.4 Tunable clustering coefficient with fixed degree distribution

In this subsection, we show how to use our model in order to generate graphs with a given degree distribution and clustering. This construction will allow us to compare graphs with a given degree distribution but with various clustering coefficients and to see the impact of clustering on the epidemic. This analysis is not possible with the random intersection graphs studied in [7]. Indeed, once the clustering coefficient and the mean degree in the graph is fixed in [7], the degree distribution is completely determined and has to be compound Poisson. In particular, the variance (as all higher moments) of the degree distribution is also fixed.

Our model has a lot more freedom in term of graphs that we can generate but still has one limitation: for a given degree distribution, there is a constraint on the maximal value of the clustering coefficient for our model. As a simple example, note that our model is not able to generate 2-regular graphs with positive clustering coefficient. In order to provide a graph with a given asymptotic degree distribution $\tilde{\mathbf{p}}$ and a positive clustering coefficient, we need the following assumptions on $\tilde{\mathbf{p}}$:

Condition 6. *We assume that the probability distribution $\tilde{\mathbf{p}}$ satisfies:*

- (i) $\sum_r r^2 \tilde{p}_r < \infty$;
- (ii) $\sum_{r \geq 3} \tilde{p}_r > 0$;
- (iii) $\tilde{p}_0 = 0$.

Under these conditions, we have the following proposition:

Proposition 7. *Let $\tilde{\mathbf{p}} = (\tilde{p}_r)_{r \geq 0}$ be a probability distribution satisfying Condition 6. We define the maximal clustering coefficient as*

$$C^{\max} := 1 - \frac{2 \sum_{r \geq 2} (r-1) \tilde{p}_r}{\sum_{r \geq 2} r(r-1) \tilde{p}_r}. \quad (2)$$

Then for any value $0 \leq C \leq C^{\max}$, there exists a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$ and a value of $\gamma \in [0, 1]$ such that the model $\tilde{G}(n, \mathbf{d}, \gamma)$ has asymptotic degree distribution $\tilde{\mathbf{p}}$ and asymptotic clustering coefficient C .

More precisely, γ is the solution of the following equation:

$$C = C(\gamma) := \frac{\sum_{r \geq 3} r(r-1)(r-2) \frac{\gamma}{(r-1)\gamma+1} \tilde{p}_r}{\sum_{r \geq 2} r(r-1) \tilde{p}_r}. \quad (3)$$

Let $F(\gamma') := \sum_{r \geq 1} \frac{r}{(r-1)\gamma'+1} \tilde{p}_r$ for all $\gamma' \in [0, 1]$, and set

$$\lambda := \begin{cases} \frac{F(\gamma)(1-\gamma)}{1-\gamma F(\gamma)} & \text{if } \gamma \neq 1, \\ \frac{1}{\sum_{r \geq 1} \tilde{p}_r/r} & \text{if } \gamma = 1. \end{cases} \quad (4)$$

Then we can define \mathbf{p} as:

$$p_r := \frac{\tilde{p}_r[(\lambda - 1)\gamma + 1]}{(r - 1)\gamma + 1} \text{ for all } r \geq 1, \text{ and } p_0 := 0. \quad (5)$$

Proof. We first show that γ , λ and \mathbf{p} are well-defined, and that \mathbf{p} is a probability distribution that satisfies Condition 1.

The solution γ of equation (3) exists and is unique. Indeed we have the following derivative for $C(\gamma)$:

$$\frac{dC}{d\gamma}(\gamma) = \frac{\sum_{r \geq 3} r(r-1)(r-2) \frac{1}{[(r-1)\gamma+1]^2} \tilde{p}_r}{\sum_{r \geq 2} r(r-1) \tilde{p}_r},$$

which is positive since $\sum_{r \geq 3} \tilde{p}_r > 0$. Hence the clustering coefficient $C(\gamma)$ is an increasing function of γ , taking values between 0 and $C(1) = C^{\max}$. Since $C \in [0, C^{\max}]$, there exists a unique $\gamma \in [0, 1]$ such that $C(\gamma) = C$.

The constant λ is well-defined: if $\gamma < 1$, then for all $r \geq 1$, $(r-1)\gamma + 1 > r\gamma$, which leads to $F(\gamma) < 1/\gamma$, i.e. $1 - \gamma F(\gamma) > 0$.

In addition, we have that $\sum_r r p_r = \lambda$. Indeed, if $\gamma = 1$, then $r p_r = \lambda \tilde{p}_r$, and summing over $r \geq 0$ gives the result. If $\gamma \neq 1$, we have that $\sum_r r p_r = [(\lambda - 1)\gamma + 1] F(\gamma) = \lambda$, where the first equality comes from (5) and the second one from (4).

We can easily verify that \mathbf{p} is a probability distribution: for all $r \geq 0$, we have that $((r-1)\gamma + 1) p_r = \tilde{p}_r[(\lambda - 1)\gamma + 1]$ (due to (5) and the fact that $\tilde{p}_0 = 0$). Summing over all $r \geq 0$ and using the fact that $\sum_r r p_r = \lambda$ finally gives that $\sum_r p_r = 1$.

We now consider the graph $\tilde{G}(n, \mathbf{d}, \gamma)$, with \mathbf{d} given by the following. For each n , let \mathbf{d} such that $|\{i : d_i = r\}| = \lfloor n p_r \rfloor$ for all $r \geq 0$. In addition we adjust the value of d_n (for instance) such that $\sum_i d_i$ is even. Then $\sum_i d_i^3 = \sum_r r^3 \lfloor n p_r \rfloor = O(n)$ due to Condition 6-(ii) and equation (5). Hence Condition 1 is satisfied by \mathbf{d} .

We can verify that $\tilde{G}(n, \mathbf{d}, \gamma)$ has asymptotic degree distribution $\tilde{\mathbf{p}}$, using Proposition 2 and equation (5), and asymptotic clustering coefficient C , using Proposition 3, equations (3) and (5). \square

A similar result can be proved with the biased clustering coefficient. In that case, $C(\gamma)$ is replaced by $C_2(\gamma) := \sum_{r \geq 3} \frac{(r-2)\gamma}{r\gamma+1-\gamma} \tilde{p}_r$ and $C_2^{\max} := \sum_{r \geq 3} \frac{r-2}{r} \tilde{p}_r$. We can see on Figure 2 that the interval of reachable clustering values is larger for the first notion of clustering. This illustration uses a power law degree distribution with exponential cutoff for the distribution $\tilde{\mathbf{p}}$: there exists a power $\tau > 0$ and a cutoff $\kappa > 0$ such that, for all $r \geq 1$, $\tilde{p}_r = c(\tau, \kappa) \cdot r^{-\tau} e^{-r/\kappa}$, where $c(\tau, \kappa) = 1/(\sum_s s^{-\tau} e^{-s/\kappa})$ is a normalizing constant. This cutoff κ allows Condition 6 to be satisfied for any power $\tau > 0$: in all figures, we will take $\kappa = 50$. In order to increase the mean degree of the graph in Figure 2, we decrease the power τ .

3 Diffusion threshold for random graphs with clustering

3.1 Diffusion model

In this section, we study a simple diffusion model depending on a single parameter $\pi \in [0, 1]$. For a given graph G , the dynamics of the diffusion is as follows: some set of nodes S starts out being active; all other nodes are inactive. When a node becomes active, each of her neighbors becomes active with probability π independently from each other. The final state of the diffusion can also be described in term of a bond percolation process in the graph G . Randomly delete each edge with probability $1 - \pi$ independently of all other edges. Denote by G_π the resulting graph. Then any node in S will activate all nodes in its connected component in G_π .

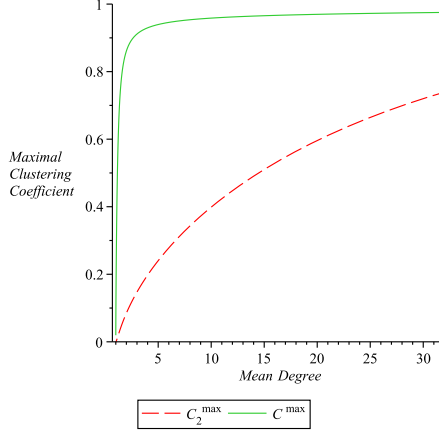


Figure 2: Reachable clustering values for both notions of clustering coefficient, with respect to the mean degree $\tilde{\lambda}$ in $\tilde{G}(n, \mathbf{d}, \gamma)$ (when the graph $\tilde{G}(n, \mathbf{d}, \gamma)$ has power law degree distribution $\tilde{p}_r \propto r^{-\tau} e^{-r/50}$)

3.2 Phase transition for the diffusion with a single activation

In this subsection we consider diffusion starting from one active node and all other nodes being inactive, and we derive conditions under which a single starting active node can activate a large fraction of the population in $G = \tilde{G}(n, \mathbf{d}, \gamma)$. This problem corresponds to the existence of a 'giant component' in the random graph obtained after bond percolation.

In order to state our result, we first need to recall some basic results about random graphs with small order. For $d \in \mathbb{N}$, let K_d be the complete graph on d vertices denoted $\{1, \dots, d\}$, with $d(d-1)/2$ edges. For $\pi \in [0, 1]$, we denote by $K_d(\pi)$ the random graph obtained from K_d after bond percolation with parameter π , i.e. each edge of K_d is kept independently of the others with probability π , otherwise it is removed.

We need to compute the probability that the component in $K_d(\pi)$ containing vertex 1 has k vertices, denoted by $f(d, k, \pi)$. Note that $f(d, d, \pi)$ is simply the probability that $K_d(\pi)$ is connected and has been computed in [10]. Indeed simple computations show that we have the recurrence relation

$$\begin{aligned} f(d, d, \pi) &= 1 - \sum_{k=1}^{d-1} \binom{d-1}{k-1} f(k, k, \pi) (1-\pi)^{k(d-k)}, \\ f(d, k, \pi) &= \binom{d-1}{k-1} f(k, k, \pi) (1-\pi)^{k(d-k)}, \end{aligned} \tag{6}$$

for any $k \leq d$.

We now define for $d \in \mathbb{N}$ and $\pi \in [0, 1]$, the random variable $\mathcal{K}(d, \pi, \gamma)$ by

$$\mathbb{P}(\mathcal{K}(d, \pi, \gamma) = k) = (1 - \gamma_d) \mathbb{1}(d = k) + \gamma_d f(d, k, \pi),$$

where f is defined in (6). In words, $\mathcal{K}(d, \pi, \gamma)$ is equal to d with probability $1 - \gamma_d$ and to the size of the component in $K_d(\pi)$ containing 1 with the remaining probability.

In addition, set for all $k \geq 1$:

$$\varrho_k := p_k(1 - \gamma_k) + \sum_{d \geq k} \frac{d}{k} f(d, k, \pi) p_d \gamma_d, \quad (7)$$

$$\varrho := \sum_{\ell} \varrho_{\ell}, \quad (8)$$

$$\mu := \sum_{\ell} \ell \varrho_{\ell} / \varrho, \quad (9)$$

$$\sigma_k := p_k(1 - \gamma_k) + \sum_{d \geq k} d f(d, k, \pi) p_d \gamma_d. \quad (10)$$

Using the notation $\tilde{\gamma}$ defined in Proposition 2, we also define (omitting the dependence on \mathbf{p} and γ):

$$\begin{aligned} L(z) &:= \sum_s \frac{\sigma_s}{\tilde{\gamma}} [1 - (1 - \pi + \pi z)^s] \\ h(z) &:= \sum_s s \frac{\varrho_s}{\varrho} (1 - \pi + \pi z)^s \\ \zeta &:= \sup\{z \in [0, 1) : \mu z(1 - \pi + \pi z) = h(z)\}. \end{aligned}$$

For a graph $G = (V, E)$ and a parameter $\pi \in [0, 1]$, we denote by $C^b(\pi)$ the size of the largest component in the bond percolated graph G_{π} .

Theorem 8. *Consider the random graph $G = \tilde{G}(n, \mathbf{d}, \gamma)$ for a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^{\infty}$, and clustering parameter $\gamma = (\gamma_r)_{r=0}^{\infty}$. Let D^* be a random variable with distribution p_r^* given by $p_{r-1}^* = \frac{r p_r}{\lambda}$ for all $r \geq 1$. We define π_c as the solution of the equation:*

$$\pi \mathbb{E}[\mathcal{K}(D^* + 1, \pi, \gamma) - 1] = 1.$$

(i) *if $\pi > \pi_c$, we have (with the notations above) that $\zeta \in (0, 1)$. In addition the asymptotic size of the largest component of the percolated graph G_{π} obtained from $\tilde{G}(n, \mathbf{d}, \gamma)$ is:*

$$C^b(\pi) / \tilde{n} \xrightarrow{p} L(\zeta) > 0.$$

(ii) *if $\pi < \pi_c$, we have $C^b(\pi) = o_p(\tilde{n})$.*

We can guess the value of the diffusion threshold π_c using a branching process approximation (Appendix A).

Note that in the particular case where $\gamma_r = 0$ for all r , we have $\mathcal{K}(d, \pi, 0) = d$ so that we get $\pi_c = \frac{\mathbb{E}[D]}{\mathbb{E}[D(D-1)]}$ where D is the typical degree in the random graph and our result reduces to a standard result in the random graphs literature (see Theorem 3.9 in [13]).

The case where a positive fraction of individuals belong to S (not only a single node) is discussed in Subsection 3.5. Now we study the effect of clustering on the diffusion with a single activation.

3.3 Effect of clustering on the diffusion for regular graphs

In this paragraph, we consider d -regular graphs ($d \geq 3$), so that $p_r = \mathbb{1}(r = d)$ for all $r \geq 0$. In this case, adding cliques does not change the asymptotic degree distribution in the graph $\tilde{G}(n, \mathbf{d}, \gamma)$, and $\tilde{p}_r = \mathbb{1}(r = d) = p_r$. In addition, we assume that $\gamma_r = \gamma$ for all $r \geq 0$ (each vertex in $G(n, \mathbf{d})$ is replaced by a clique with probability γ). We are interested in the effect of clustering on the diffusion threshold π_c on the one hand, and its effect on the epidemic size on the other hand.

We have the following result for the diffusion threshold in regular graphs:

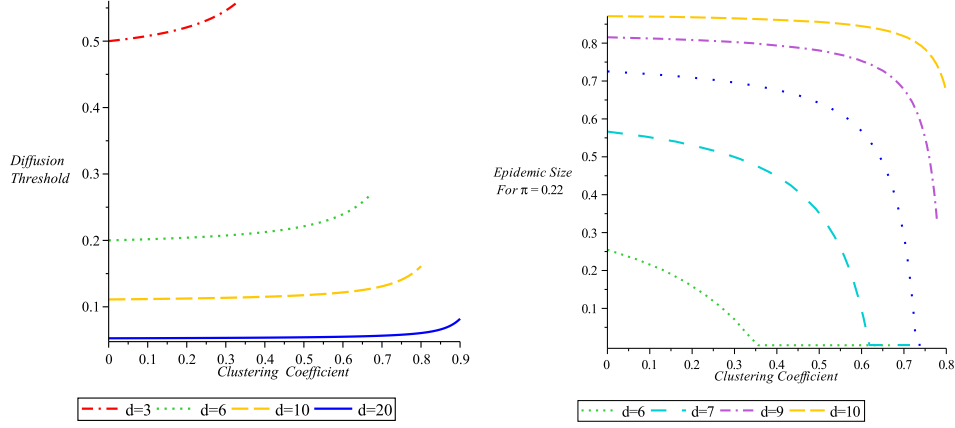


Figure 3: On the left: Evolution of the diffusion threshold with respect to the clustering coefficient for d -regular graphs. On the right: Evolution of the epidemic size with respect to the clustering coefficient for d -regular graphs (with infection probability $\pi = 0.22$).

Proposition 9. Let $d \geq 3$. We consider the asymptotic degree distribution $\tilde{\mathbf{p}} = (\mathbb{1}_{r=d})_{r \geq 0}$. Let $0 \leq C^{(1)} < C^{(2)} \leq C^{\max} = 1 - 2/d$. For each $j = 1, 2$, let $(\mathbf{d}_j, \mathbf{p}_j, \gamma_j)$ be chosen according to Proposition 7, such that $\tilde{G}^{(j)} = \tilde{G}(n, \mathbf{d}_j, \gamma_j)$ has asymptotic degree sequence $\tilde{\mathbf{p}}$ and asymptotic clustering coefficient $C^{(j)}$. Let $\pi_c^{(j)}$ be the diffusion threshold defined in Theorem 8 for the random graph $\tilde{G}^{(j)}$, $j = 1, 2$. Then we have:

$$\pi_c^{(1)} \leq \pi_c^{(2)}$$

In our graph model, the diffusion threshold for a random d -regular graph increases as the clustering coefficient increases.

Proof. Both notions of clustering coefficient considered in 2.3 are the same for random d -regular graphs, and we have that: $C(\gamma) = \frac{d-2}{d-1+1/\gamma}$. In particular, $C^{(1)} < C^{(2)}$ implies that $\gamma_1 < \gamma_2$.

According to Theorem 8, the diffusion threshold $\pi_c^{(j)}$, $j = 1, 2$, is the solution of the following equation:

$$\pi \mathbb{E}[\mathcal{K}(d, \pi, \gamma_j) - 1] = 1. \quad (11)$$

Using the definition of \mathcal{K} , equation (11) becomes $\pi = F(\pi, \gamma_j)$, where

$$F : \begin{cases} [0, 1]^2 & \longrightarrow [0, 1] \\ (\pi, \gamma) & \mapsto \frac{1}{d-1+\gamma(\sum_{k=1}^d kf(d, k, \pi) - d)}. \end{cases}$$

Let $\pi \in [0, 1]$. Then $F(\pi, \gamma_1) \leq F(\pi, \gamma_2)$ since $\gamma_1 < \gamma_2$ and $\sum_{k=1}^d kf(d, k, \pi) \leq d$. Indeed, $f(d, k, \pi)$, $k \leq d$, is the probability that the connected component of a vertex inside $K_d(\pi)$ has k vertices in $K_d(\pi)$. Hence $\sum_{k=1}^d kf(d, k, \pi)$ is the mean size of that component inside $K_d(\pi)$, so that $\sum_{k=1}^d kf(d, k, \pi) \leq d$.

Therefore the curve Γ_2 of $\pi \mapsto F(\pi, \gamma_2)$ is above the curve Γ_1 of $\pi \mapsto F(\pi, \gamma_1)$. Let A_1 (resp. A_2) be the intersection between Γ_1 (resp. Γ_2) and the first bisector. Both functions are continuous, thus the first coordinate of A_1 is less than or equal to the one of A_2 , that is to say $\pi_c^{(1)} \leq \pi_c^{(2)}$. \square

More precisely, Figure 3 (on the left) shows how the diffusion threshold increases with the clustering coefficient, for different values of d : in other words, clustering decreases the range of π ($\pi \in (\pi_c; 1]$) for which a single individual can turn a positive proportion of the population into infected individuals.

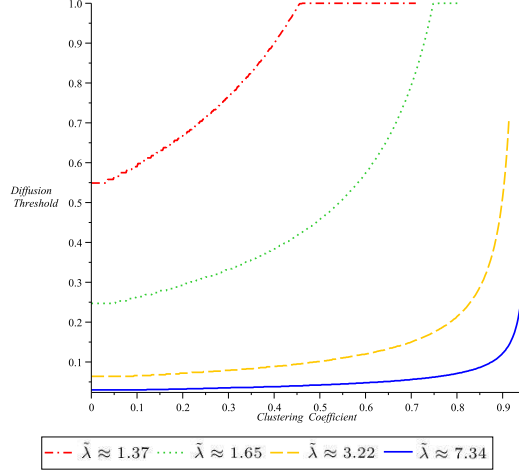


Figure 4: Evolution of the diffusion threshold with respect to the clustering coefficient in a graph with mean degree $\tilde{\lambda}$, with respect to the clustering coefficient C (for a fixed power law degree distribution).

In addition, the epidemic size also decreases with the clustering: in Figure 3 (on the right), we plot the ratio of the largest connected component in the percolated graph over the whole population. When the starting infected individual is a vertex chosen uniformly at random, this ratio also corresponds to the probability of explosion. Hence, as the clustering increases, it 'inhibits' the diffusion process. These results are in accordance with [11].

These results are intuitive (for d -regular graphs) in the sense that the removal of edges inside cliques can stop the diffusion *inside* a clique in the graph $\tilde{G}(n, \mathbf{d}, \gamma)$, while this phenomenon does not occur in the original graph $G(n, \mathbf{d})$.

3.4 Effect of clustering on the diffusion for graphs with power law degree distribution

In Figure 4, we consider a power law degree distribution with exponential cutoff: $\tilde{p}_r \propto r^{-\tau} e^{-r/50}$, with parameters $\tau = 2.9$, $\tau = 2.5$, $\tau = 1.81$, $\tau = 1.3$, so that the mean degree is respectively $\tilde{\lambda} \approx 1.37$, $\tilde{\lambda} \approx 1.65$, $\tilde{\lambda} \approx 3.22$, $\tilde{\lambda} \approx 7.3$. We plot the diffusion threshold π_c for the graph given by Proposition 7, when the degree distribution is $\tilde{\mathbf{p}}$ and the clustering coefficient varies from 0 to C^{\max} . We observe the same phenomenon as the one we proved for d -regular graphs, *i.e.* clustering decreases the range of π ($\pi \in (\pi_c; 1]$) for which a single individual can turn a positive proportion of the population into infected individuals.

3.5 Phase transition for the diffusion with degree based activation

In this subsection, we allow a positive fraction of nodes to be active at the beginning of the diffusion process. More precisely, on a given graph G , the set S of initial active nodes is random, and each node of degree d in G belongs to S with some probability $\alpha_d > 0$, independently for each node. We set $\boldsymbol{\alpha} = (\alpha_d)_{d \geq 0}$.

Using the notation $\tilde{\gamma}$ defined in Proposition 2, and definitions (7) to (9), we define (omitting the dependence on $\boldsymbol{\alpha}$, \mathbf{p} and γ):

$$\begin{aligned}
L(z) &:= \sum_s \frac{(1-\gamma_s)p_s}{\tilde{\gamma}} [1 - (1-\alpha_s)(1-\pi+\pi\zeta)^s] \\
&\quad + \sum_{d \geq s} df(d, s, \pi) \frac{\gamma_d p_d}{\tilde{\gamma}} [1 - (1-\alpha_d)^s (1-\pi+\pi\zeta)^s] \\
h(z) &:= \sum_s s(1-\gamma_s)p_s [1 - (1-\alpha_s)(1-\pi+\pi\zeta)^s] / \varrho \\
&\quad + \sum_{d \geq s} df(d, s, \pi) \gamma_d p_d [1 - (1-\alpha_d)^s (1-\pi+\pi\zeta)^s] / \varrho \\
\zeta &:= \sup\{z \in [0, 1] : \mu z(1-\pi+\pi z) = h(z)\}.
\end{aligned} \tag{12}$$

Theorem 10. Consider the random graph $G = \tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$ for a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$, and clustering parameter $\boldsymbol{\gamma} = (\gamma_r)_{r=0}^\infty$. We are given an activation set S drawn according to the distribution $\boldsymbol{\alpha}$. Then we have, for the diffusion model defined in 3.1: if $\zeta = 0$, or if $\zeta \in (0, 1]$, and further ζ is such that there exists $\varepsilon > 0$ with $\lambda z(1-\pi+\pi z) < h(z)$ for $z \in (\zeta - \varepsilon, \zeta)$, then we have that the size $C^b(\pi, \boldsymbol{\alpha})$ of the active nodes at the end of the diffusion verifies:

$$C^b(\pi, \boldsymbol{\alpha}) / \tilde{n} \xrightarrow{P} L(\zeta).$$

Heuristically, taking $\alpha_s = 0$ for all s in the definitions of the previous theorem allows to recover the result of Theorem 8.

4 Symmetric threshold model for random graphs with clustering

4.1 Symmetric threshold model

We now describe the symmetric threshold model on a finite graph $G = (V, E)$, with given thresholds $k(v)$, for $v \in V$. The progressive dynamics of the epidemic on the finite graph G operates as follows: some set of nodes S starts out being active; all other nodes are inactive. Time operates in discrete steps $t = 1, 2, 3, \dots$. At a given time t , any inactive node v becomes active if its number of active neighbors is at least $k(v) + 1$. This in turn may cause other nodes to become active. It is easy to see that the final set of active nodes (after n time steps if the network is of size n) only depends on the initial set S (and not on the order of the activations) and can be obtained as follows: set $Y_v = \mathbf{1}(v \in S)$ for all v . Then as long as there exists v such that $\sum_{w \sim v} Y_w > k(v)$, set $Y_v = 1$, where $w \sim v$ means that v and w share an edge in G . When this algorithm finishes, the final state of node v is represented by Y_v : $Y_v = 1$ if node v is active and $Y_v = 0$ otherwise. In this paper, we do not analyze the dynamics of the epidemics and concentrate on the final state only.

We allow the threshold $k(v)$ of a node v to be a random variable with distribution depending on the degree of v , and such that thresholds are independent among nodes. More precisely, for each $s \geq 0$, let $(t_{s\ell})_{0 \leq \ell \leq s}$ be a probability distribution. We draw independent thresholds $k(i)$, for $i \in V$. Knowing that the degree d_i of node i is s , threshold $k(i)$ is drawn according to the conditional probability distribution $(t_{s\ell})_{0 \leq \ell \leq s}$: $\mathbb{P}(k(i) = \ell | d_i = s) = t_{s\ell}$. We say that random thresholds $\mathbf{k} = (k(i))_{i \in V}$ are drawn according to $\mathbf{t} = (t_{s\ell})_{s, \ell}$.

To simplify, we define an adaptation of the symmetric threshold model for the random graph $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$: we draw random thresholds \mathbf{k} for each vertex i in the original graph $G(n, \mathbf{d})$. When a vertex i of $G(n, \mathbf{d})$ is replaced by a clique in $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$, we associate to each vertex inside the clique the original threshold $k(i)$, so that vertices inside a clique have the same threshold (also referred to as the “threshold of the clique”). We still denote by $k(v)$ the threshold of a vertex v in $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$.

4.2 Phase transition for the symmetric threshold model with a single activation

We show that there is a phase transition for a single active node to turn a positive fraction of the population into active nodes. For a graph $G = (V, E)$ and thresholds $\mathbf{k} = (k(v))_{v \in V}$, we consider the largest connected component of the induced subgraph in which we keep only vertices of threshold zero. We call the vertices in this component pivotal players: if only one pivotal player becomes active then the whole set of pivotal players will eventually become active. In particular, if the set of pivotal players is large (*i.e.* of order $\Theta(|V|)$, as $|V| \rightarrow \infty$), then a single pivotal player u can trigger a global cascade (*i.e.* the number of active nodes at the end of the epidemic starting from u is of order $\Theta(|V|)$).

We consider the random graph $G = \tilde{G}(n, \mathbf{d}, \gamma)$, and random thresholds drawn according to the distribution \mathbf{t} . For a node v , we denote by $C(v, \mathbf{t})$ the final number of active vertices, when the initial state consists of only v active and all other nodes are inactive. Informally, we say that $C(v, \mathbf{t})$ is the size of the cascade induced by node v ; if $C(v, \mathbf{t}) = \Theta_p(\tilde{n})$, we say that node v can trigger a global cascade.

Using the notation $\tilde{\gamma}$ defined in Proposition 2 and the binomial probabilities $b_{sr}(p)$ defined at the end of Section 1, we set (omitting the dependence on \mathbf{t} , \mathbf{p} and γ):

$$\begin{aligned} L(z) &:= \sum_s \frac{[s\gamma_s + (1 - \gamma_s)] p_s}{\tilde{\gamma}} t_{s0} (1 - z^s) \\ &\quad + \sum_s \frac{(1 - \gamma_s) p_s}{\tilde{\gamma}} \left(1 - t_{s0} - \sum_{\ell \neq 0} t_{s\ell} \sum_{r \geq s-\ell} b_{sr}(\zeta) \right), \\ h(z) &:= \sum_s s p_s [t_{s0} z^s + \gamma_s (1 - t_{s0}) z] \\ &\quad + \sum_s p_s (1 - \gamma_s) \sum_{s \geq \ell \neq 0} t_{s\ell} \sum_{r \geq s-\ell} r b_{sr}(z), \\ \zeta &:= \sup\{z \in [0, 1) : \lambda z^2 = h(z)\}. \end{aligned} \tag{13}$$

Theorem 11. *Consider the random graph $\tilde{G}(n, \mathbf{d}, \gamma)$ for a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$, and clustering parameter $\gamma = (\gamma_r)_{r=0}^\infty$. Let \mathbf{t} be a family of probability distributions, and \mathbf{k} random thresholds drawn according to \mathbf{t} in the original graph $G(n, \mathbf{d})$ (*i.e.* if i is a vertex in $G(n, \mathbf{d})$ replaced by a clique in $\tilde{G}(n, \mathbf{d}, \gamma)$, then all vertices in the clique have the same threshold $k(i)$). We call the following condition the cascade condition:*

$$\sum_r r(r-1)p_r t_{r0} > \sum_r r p_r. \tag{14}$$

Let $\tilde{\mathcal{P}}^{(n)}$ be the set of pivotal players in $\tilde{G}(n, \mathbf{d}, \gamma)$.

(i) *If the cascade condition (14) is satisfied, then there is a unique $\xi \in (0, 1)$ such that*

$$\sum_d d p_d t_{d0} (1 - \xi^{d-1}) = \lambda (1 - \xi) \tag{15}$$

and we have:

$$\frac{|\tilde{\mathcal{P}}^{(n)}|}{\tilde{n}} \xrightarrow{p} \sum_d \frac{[d\gamma_d + (1 - \gamma_d)] p_d t_{d0}}{\tilde{\gamma}} (1 - \xi^d) > 0, \tag{16}$$

where $\tilde{\gamma}$ is defined in Proposition 2. Moreover, for any $u \in \tilde{\mathcal{P}}^{(n)}$, we have whp

$$\liminf \frac{C(u, \mathbf{t})}{\tilde{n}} \geq L(\zeta) > 0 \tag{17}$$

where ζ is defined by (13). If in addition $\zeta = 0$ or ζ is such that there exists $\varepsilon > 0$ with $\lambda z^2 < h(z)$ for $z \in (\zeta - \varepsilon, \zeta)$, then we have for any $u \in \tilde{\mathcal{P}}^{(n)}$:

$$C(u, \mathbf{t})/\tilde{n} \xrightarrow{p} L(\zeta). \quad (18)$$

(ii) If $\sum_r r(r-1)p_r t_{r0} < \sum_r r p_r$, for a uniformly chosen player u , we have $C(u, \mathbf{t}) = o_p(\tilde{n})$. The same result holds if $o(n)$ players are chosen uniformly at random.

When $\gamma_r = 0$ for all $r \geq 0$, Theorem 11 corresponds to the result of [20]. When we add cliques in the graph, the effect on the epidemic can be described by the following lemma.

Lemma 12. *We consider a clique in $\tilde{G}(n, \mathbf{d}, \gamma)$ where all vertices are inactive, and at least one of them has a neighbor outside the clique which is active. If the threshold k of the (vertices in the) clique is zero, then the epidemic will propagate to the whole clique. On the contrary, if k is positive, then the clique cannot become active, even if all neighbors outside are active.*

Indeed, if $k = 0$, each vertex in the clique needs only one active neighbor to become active. If $k > 0$, each vertex in the clique needs at least two active neighbors to become active. Yet each vertex of the clique has only one (active) neighbor outside, other neighbors being (inactive ones) inside the clique.

Hence a clique with positive threshold in which all vertices are initially inactive will always stop the epidemic. This simple observation allows to make a comparison between the epidemic in the original graph $G(n, \mathbf{d})$ and the epidemic in the graph $\tilde{G}(n, \mathbf{d}, \gamma)$ with additional cliques: since cliques have a tendency to stop epidemic, it is also easy to see that if there is no global cascade in $G(n, \mathbf{d})$, then there is no one in $\tilde{G}(n, \mathbf{d}, \gamma)$ (more details are given in the proof, 5.5).

The fact that the converse is also true is more remarkable, since the cliques with positive threshold stop the epidemic. In fact, those cliques will reduce the *size* of the cascade, but they have no impact on the fact that cascade is possible or not. Indeed it is shown in [20] that a global cascade is possible in the graph $G(n, \mathbf{d})$ if and only if the set of pivotal players (in $G(n, \mathbf{d})$) is large. Note that one direction of this equivalence is easy, and holds for any graph (in particular, this is still true for $\tilde{G}(n, \mathbf{d}, \gamma)$): if the set of pivotal players is large, any pivotal player which is initially active can trigger a global cascade, as explained at the beginning of the subsection. Let us assume now that there is a global cascade in $G(n, \mathbf{d})$. Using the equivalence shown in [20], the set of pivotal players (in $G(n, \mathbf{d})$) is large. This implies that the set of pivotal players in the graph $\tilde{G}(n, \mathbf{d}, \gamma)$ (with additional cliques) is also large (details are given in the proof, 5.5), so that there is a cascade in the graph $\tilde{G}(n, \mathbf{d}, \gamma)$.

Hence there is a global cascade in $\tilde{G}(n, \mathbf{d}, \gamma)$ if and only if there is one in the original graph $G(n, \mathbf{d})$. This explains why the cascade condition (14) only depends on the original distribution \mathbf{p} and threshold distribution \mathbf{t} (and not on γ). Yet these two graphs ($G(n, \mathbf{d})$ and $\tilde{G}(n, \mathbf{d}, \gamma)$) have not the same asymptotic degree distribution. What is interesting now is to compare two graphs that have the same asymptotic degree distribution $\tilde{\mathbf{p}} = (\tilde{p}_r)_{r \geq 0}$, but different clustering coefficients.

4.3 Effect of clustering on the contagion threshold

We use our results to highlight the effect of clustering for the game-theoretic contagion model proposed by Blume [4] and Morris [22]. Consider a graph G in which the nodes are the individuals in the population and there is an edge (i, j) if i and j can interact with each other. Each node has a choice between two possible actions labeled A and B . On each edge (i, j) , there is an incentive for i and j to have their actions match, which is modeled as the following coordination game parametrized by a real number $q \in (0, 1)$: if i and j choose A (resp. B), they each receive a payoff of q (resp. $(1 - q)$); if they choose opposite actions, then they receive a payoff of 0. Then the total payoff of a player is the sum of the payoffs with each of her neighbors. If the degree of node i is d_i and S_i^B is the number of its neighbors playing B , then the payoff to i from choosing A is $q(d_i - S_i^B)$ while the payoff from choosing B is $(1 - q)S_i^B$. Hence, in a

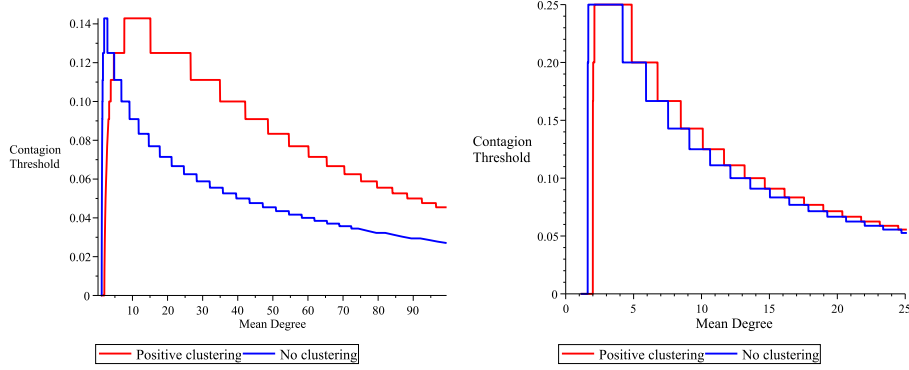


Figure 5: On the left: Contagion thresholds in two graphs with the same degree distribution $\tilde{p}_r \propto r^{-\tau} e^{-r/50}$. On the right: Contagion thresholds in two graphs with the same degree distribution $\tilde{p}_r = e^{-\lambda} \lambda^{r-1} / (r-1)!$.

best-response dynamic, i should adopt B if $S_i^B > qd_i$ and A if $S_i^B \leq qd_i$. A number of qualitative insights can be derived from such a model even at this level of simplicity [18]. Specifically, consider a network where all nodes initially play A . If a small number of nodes are forced to adopt strategy B (the seed) and we apply best-response updates to other nodes in the network, then these nodes will be repeatedly applying the following rule: switch to B if enough of your neighbors have already adopted B . There can be a cascading sequence of nodes switching to B such that a network-wide equilibrium is reached in the limit. Note that the dynamics of the contagion process is deterministic once the seed is fixed as opposed to the diffusion process. The contagion process is a particular case of the symmetric threshold model. Indeed the threshold distribution is given by: $t_{s\ell} = \mathbb{1}(\lfloor qs \rfloor = \ell)$ for all $0 \leq \ell \leq s$. The cascade condition (14) is satisfied if and only if the parameter q of the contagion is greater than the contagion threshold

$$q_c := \sup \left\{ q : \sum_{r < q^{-1}} r(r-1)p_r > \sum_r r p_r \right\}. \quad (19)$$

We restrict ourselves to the case where $\gamma_r = \gamma$ for all $r \geq 0$ and we use Proposition 7 to construct two graphs with the same asymptotic degree distribution $\tilde{\mathbf{p}}$, one with a positive clustering coefficient, the other with no clustering. We then compare the contagion thresholds in these two graphs.

In Figure 5 on the left, we consider a power law degree distribution with parameter $\tau > 0$ and exponential cutoff: for all $r \geq 1$, $\tilde{p}_r \propto r^{-\tau} e^{-r/50}$. On the one hand, we consider (in red) the graph $G^1(\tau)$ for $C = C^{\max}$ (so that $\gamma = 1$ and the distribution in the original graph is: $p_r \propto r^{-(\tau+1)} e^{-r/50}$). On the other hand, we consider (in blue) the graph $G^0(\tau)$ given by Proposition 7 for $C = 0$ (so that $\gamma = 0$ and the distribution in the original graph is: $p_r = \tilde{p}_r$). In Figure 5 (on the left), we make the parameter τ vary: the red (resp. blue) curve corresponds to the contagion threshold $q_c^1(\tau)$ (resp. $q_c^0(\tau)$) of the graph $G^1(\tau)$ (resp. $G^0(\tau)$), defined in (19). Contagion thresholds are given with respect to the mean degree $\tilde{\lambda} = \sum_r r \tilde{p}_r$ (that is a decreasing function of τ).

In Figure 5 on the right, we consider another form for the degree distribution $\tilde{\mathbf{p}}$: let $\lambda > 0$, and set $\tilde{p}_r = e^{-\lambda} \lambda^{r-1} / (r-1)!$ for all $r \geq 1$. As before, we consider the graph $G^1(\lambda)$ for $C = C^{\max}$ ($\gamma = 1$ and \mathbf{p} is a Poisson distribution with parameter λ : $p_r = e^{-\lambda} \lambda^r / r!$), and the graph $G^0(\lambda)$ given by Proposition 7 for $C = 0$ ($\gamma = 0$ and $p_r = \tilde{p}_r$). In Figure 5 (on the right), we plot the contagion thresholds for these two graphs, with respect to the mean degree $\tilde{\lambda} = \lambda + 1$.

Both left and right-hand sides of Figure 5 show that, when the mean degree $\tilde{\lambda}$ of the graph is low, the contagion threshold q_c^0 of the graph with no clustering is greater than the threshold q_c^1 of the graph with positive clustering. Hence, if the parameter q of the contagion process is in the interval $]q_c^1, q_c^0[$, a

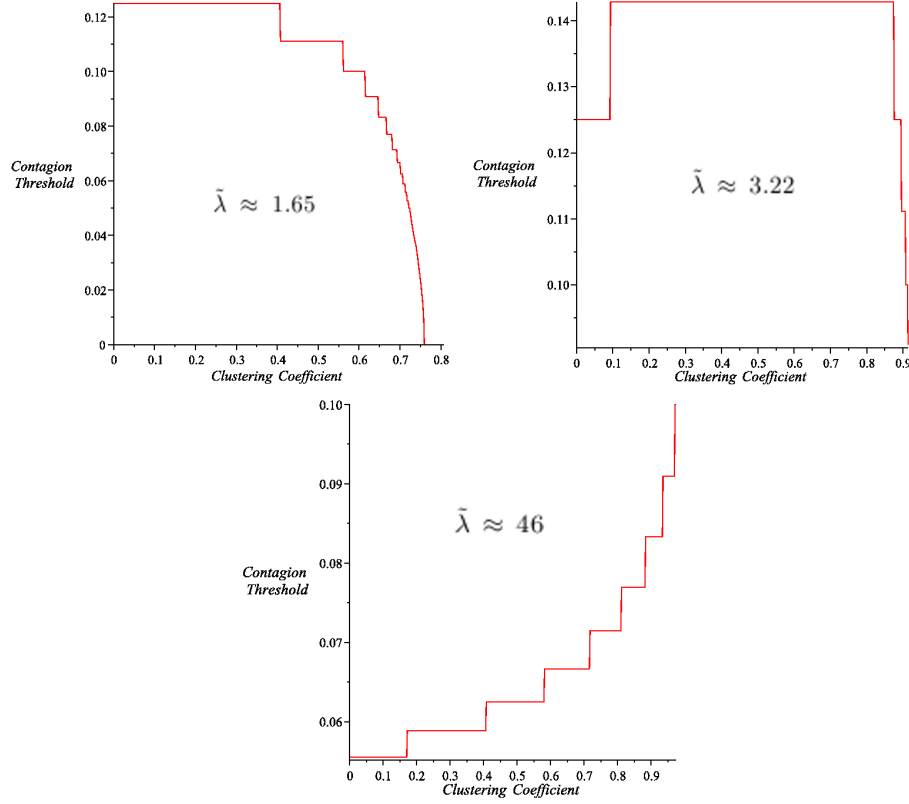


Figure 6: Evolution of the contagion threshold in a graph with mean degree $\tilde{\lambda}$ with respect to the clustering coefficient C (for a fixed power law degree distribution).

global cascade is possible only in the graph with no clustering: in that case, the clustering 'inhibits' the contagion process. On the contrary, for high values of the mean degree, we have that $q_c^0 < q_c^1$, so the clustering increases the range of parameter q for which a global cascade is possible.

Let us fix the parameter $q \in (0, 1)$ of the contagion (q sufficiently low): this corresponds to a horizontal cut in Figure 5. The interval of mean degrees $\tilde{\lambda}$ for which a global cascade is possible moves to the right when the clustering increases. Hence when the parameter of the contagion is fixed, clustering favors contagion processes on graphs with a higher mean degree.

Now we study more precisely what happens if we fix the mean degree in the graph (which corresponds to a vertical cut in Figure 5), and increase the clustering coefficient between 0 and its maximal value C^{\max} (even if we consider the first notion of clustering coefficient, the same phenomenon appears with the second notion).

In Figure 6 on the top left corner (resp. top right corner, bottom), we consider a power law degree distribution with exponential cutoff: $\tilde{p}_r \propto r^{-\tau} e^{-r/50}$, with parameter $\tau = 2.5$ (resp. $\tau = 1.81$, $\tau = 0.1$). We plot the contagion threshold q_c for the graph given by Proposition 7, when the degree distribution is $\tilde{\mathbf{p}}$ and the clustering coefficient varies from 0 to C^{\max} . We consider three different slices of Figure 5 (left), and we go from the blue curve ($C = 0$) to the red one ($C = C^{\max}$), progressively increasing the clustering coefficient. For a very low value of the mean degree ($\tilde{\lambda} \approx 1.65$, top left corner of Figure 6), the contagion threshold decreases with the clustering. The opposite happens when the mean degree is very high ($\tilde{\lambda} \approx 46$, bottom). In addition, for some intermediate values of the mean degree, as for $\tilde{\lambda} \approx 3.22$ (top right corner), low values of the clustering 'helps' the contagion process, but, as the clustering coefficient becomes higher, the opposite happens: it 'inhibits' more and more the contagion process.

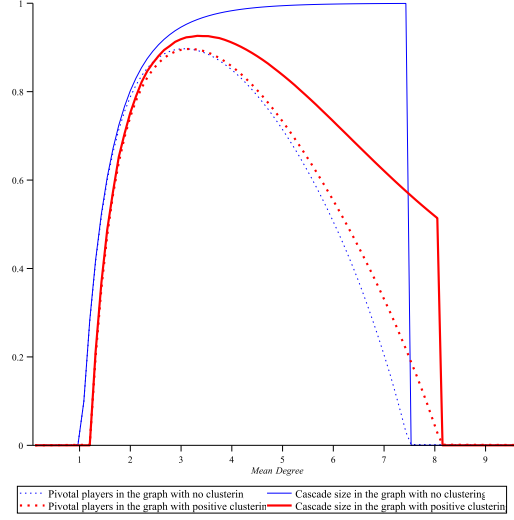


Figure 7: Set of pivotal players and cascade sizes for $q = 0.15$

We see that the impact of clustering is different for low values of the mean degree and for high values of the mean degree. In the low values regime for the mean degree, the contagion is more and more difficult, as the clustering increases. On the contrary, in the high values regime, the higher the clustering is, the more it 'helps' the contagion. When the value of the mean degree is exactly between these two cases, the effect of clustering is ambiguous: a low clustering coefficient 'helps' the contagion process, but a high one 'inhibits' the process.

4.4 Effect of clustering on the cascade size for the contagion model

We still consider the game-theoretic contagion model proposed by Morris [22] (described in the introduction), and the case where $\gamma_r = \gamma$ for all $r \geq 0$. In this subsection, the parameter $q \in (0, 1)$ of the contagion process is fixed, and we want to highlight the effect of the clustering on the cascade size.

First we compare two graphs with the same asymptotic degree distribution $\tilde{\mathbf{p}}$, one having a positive clustering coefficient, the other having no clustering. In Figure 7, we plot the sizes of the cascade and the pivotal players set for each of these graphs.

More precisely, in Figure 7, we fix $q = 0.15$. The red curves correspond to a graph with positive clustering, constructed as follows: we start from a Poisson distribution with parameter λ for \mathbf{p} , and $\gamma = 0.2$. This gives $\tilde{p}_r = \frac{0.2r+0.8}{0.2\lambda+0.8} \frac{e^{-\lambda}\lambda^r}{r!}$ and clustering coefficient $C = \frac{0.2\lambda}{0.2\lambda+1.2} > 0$. The blue curves correspond to a graph with the same asymptotic distribution $\tilde{\mathbf{p}}$, but no clustering (in that case, $\mathbf{p} = \tilde{\mathbf{p}}$ and $\gamma = 0$). We make the parameter λ vary, and the sizes of the cascade (solid lines) and the pivotal players set (dot lines) are plot with respect to the mean degree $\tilde{\lambda} = \sum_r r\tilde{p}_r$ in the graph.

For each graph, we observe that there is a cascade if and only if the set of pivotal players is large, as explained in 4.2. In addition, the interval of mean degrees $\tilde{\lambda}$ for which a cascade is possible moves to the right when the clustering coefficient increases, which is consistent with our observations on Figure 5. Finally, we observe that the size of the cascade (when it exists) decreases with the clustering. This comes from the fact that cliques of degree $d \geq q^{-1}$ (*i.e.* cliques with positive threshold) stop the contagion process (as explained in Lemma 12). In the extremal case when $\gamma = 1$ (each vertex of degree d is replaced by a clique of size d), the cascade is exactly the set of pivotal players. When the probability γ of replacing a vertex by a clique increases, the cascade triggered by a pivotal player becomes closer and closer to the set of pivotal players only (until it is exactly the set of pivotal players). This observation is confirmed in

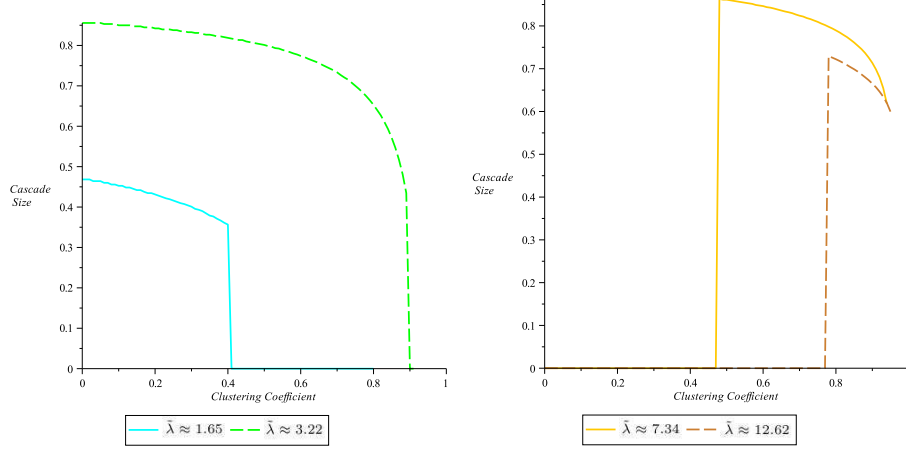


Figure 8: Effect of the clustering on the cascade size for $q = 0.12$, in a graph with a fixed power law degree distribution of mean $\tilde{\lambda}$.

Figure 8.

To study more precisely the effect of clustering on the cascade size, we plot (in Figure 8) the cascade size for $q = 0.12$, with respect to the clustering coefficient of a graph with power law degree distribution with exponential cutoff: $\tilde{p}_r \propto r^{-\tau} e^{-r/50}$, with parameter $\tau = 2.5$ (resp. $\tau = 1.81, \tau = 1.3, \tau = 1$), so that the mean degree of the graph is $\tilde{\lambda} \approx 1.65$ (resp. $\tilde{\lambda} \approx 3.22, \tilde{\lambda} \approx 7.34, \tilde{\lambda} \approx 12.62$). Note that the values of the mean degree $\tilde{\lambda}$ on the right-hand side of Figure 8 correspond to the case where the clustering 'helps' the contagion to spread, while the case $\tilde{\lambda} \approx 1.65$ corresponds to the case where the clustering 'inhibits' the contagion, as detailed in the previous subsection. As for Figure 7, we observe that the cascade size decreases with the clustering coefficient, when the cascade size is positive (*i.e.* when a cascade is possible). The fact that a cascade is not possible for low values of clustering (right-hand side) comes from the fact that, for a fixed parameter q , the interval of $\tilde{\lambda}$ for which a cascade is possible moves to the right, as observed in Figures 5 and 7.

4.5 Phase transition for the symmetric threshold model with degree based activation

In this subsection, we allow a positive fraction of nodes to be active at the beginning of the diffusion process. More precisely, on a given graph G , the set S of initial active nodes is random, and each node of degree d in G belongs to S with some probability $\alpha_d > 0$, independently for each node. We set $\alpha = (\alpha_d)_{d \geq 0}$.

We define an adaptation of the usual degree based activation for the random graph $\tilde{G}(n, \mathbf{d}, \gamma)$ (so that the initial activation differs from the one in Subsection 3.5). First we draw independent random variables for each vertex in the original graph $G(n, \mathbf{d})$. More precisely, for each vertex i (of degree d_i in $G(n, \mathbf{d})$), we draw a Bernoulli random variable $a(i)$ with parameter α_{d_i} . When a vertex i of $G(n, \mathbf{d})$ is replaced by a clique in $\tilde{G}(n, \mathbf{d}, \gamma)$, we associate to each vertex inside the clique the same activation variable $a(i)$ (if i is not replaced by a clique, it keeps its own activation variable). Each vertex v in $\tilde{G}(n, \mathbf{d}, \gamma)$ belongs to the initial seed S if and only if $a(v) = 1$. Note that each node of degree d in $\tilde{G}(n, \mathbf{d}, \gamma)$ belongs to S with probability $\alpha_d > 0$ (since vertices inside the clique generated by i have the same degree as i). Thus the only difference with the usual degree based activation is that activation variables are not independent inside a clique: either the whole clique belongs to the initial seed S , either no vertex in the clique belongs to S .

Using the notation $\tilde{\gamma}$ defined in Proposition 2 and the binomial probabilities $b_{sr}(p)$ defined at the end of Section 1, we define (omitting the dependence on $\alpha, \mathbf{t}, \mathbf{p}$ and γ):

$$\begin{aligned}
L(z) &:= \sum_s \frac{[s\gamma_s + (1 - \gamma_s)]p_s}{\tilde{\gamma}} [(1 - \alpha_s)t_{s0}(1 - z^s) + \alpha_s] \\
&\quad + \sum_s \frac{(1 - \gamma_s)p_s}{\tilde{\gamma}} (1 - \alpha_s) \left(1 - t_{s0} - \sum_{\ell \neq 0} t_{s\ell} \sum_{r \geq s-\ell} b_{sr}(z) \right), \\
h(z) &:= \sum_s (1 - \alpha_s)sp_s [t_{s0}z^s + \gamma_s(1 - t_{s0})z] \\
&\quad + \sum_s (1 - \alpha_s)p_s(1 - \gamma_s) \sum_{s \geq \ell \neq 0} t_{s\ell} \sum_{r \geq s-\ell} rb_{sr}(z), \\
\zeta &:= \sup\{z \in [0, 1) : \lambda z^2 = h(z)\}.
\end{aligned} \tag{20}$$

Theorem 13. Consider the random graph $\tilde{G}(n, \mathbf{d}, \gamma)$ for a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$, and clustering parameter $\gamma = (\gamma_r)_{r=0}^\infty$. Let \mathbf{t} be a family of probability distributions, and \mathbf{k} random thresholds drawn according to \mathbf{t} in the original graph $G(n, \mathbf{d})$ (i.e. if i is a vertex in $G(n, \mathbf{d})$ replaced by a clique in $\tilde{G}(n, \mathbf{d}, \gamma)$, then all vertices in the clique have the same threshold $k(i)$). We are given an activation set S drawn according to the distribution α (so that vertices in the same clique are either all active or all inactive). Then we have, for the symmetric threshold model defined in 4.1: if $\zeta = 0$, or if $\zeta \in (0, 1]$, and further ζ is such that there exists $\varepsilon > 0$ with $\lambda z^2 < h(z)$ for $z \in (\zeta - \varepsilon, \zeta)$, then we have that the size $C(\mathbf{t}, \alpha)$ of the active nodes at the end of the symmetric threshold process verifies:

$$C(\mathbf{t}, \alpha)/\tilde{n} \xrightarrow{p} L(\zeta).$$

Heuristically, taking $\alpha_s = 0$ for all s in the definitions of the previous theorem allows to recover the result of Theorem 11. When $\gamma_r = 0$ for all $r \geq 0$, we recover a result in [20].

If we apply this result to the case where thresholds are constant among nodes (i.e. there exists an integer k such that $k(v) = k$ for each vertex v), our model corresponds to a slight modification of the usual bootstrap percolation. Indeed the initial activation here is not independent among nodes that belong to the same clique.

5 Proofs

In the whole section, we consider a sequence \mathbf{d} satisfying Condition 1 with probability distribution $\mathbf{p} = (p_r)_{r=0}^\infty$.

5.1 Configuration Model

In order to prove Theorem 8, it will be more convenient to work with the configuration model $G^*(n, \mathbf{d})$ (see for instance [5]): each vertex i , $1 \leq i \leq n$, has d_i half-edges, and the random graph $G^*(n, \mathbf{d})$ is obtained by taking a uniform matching among all possible matchings of half-edges into pairs. Conditioned on this multigraph being simple, it is distributed as $G(n, \mathbf{d})$. Condition 1 implies (in particular) that

$$\liminf \mathbb{P}(G^*(n, \mathbf{d}) \text{ is simple}) > 0 \tag{21}$$

(see [14]), which allows to transfer directly results that hold in probability for $G^*(n, \mathbf{d})$ to the model $G(n, \mathbf{d})$.

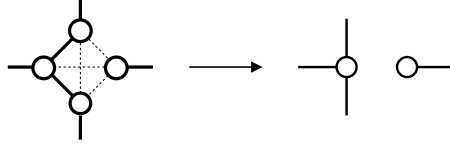


Figure 9: Transformation of a subgraph in $\tilde{G}^*(n, \mathbf{d}, \gamma)$ (or $\tilde{G}(n, \mathbf{d}, \gamma)$) by ϕ

As for the simple graph, we consider the model $\tilde{G}^*(n, \mathbf{d}, \gamma)$: we associate to each $i \in \{1, \dots, n\}$ a Bernoulli variable $X(i)$ with parameter γ_{d_i} , all variables being independent. If $X(i) = 1$, we replace node i by a clique of size d_i in which each vertex has exactly $d_i - 1$ neighbors inside the clique, and one half-edge outside. Then we match half-edges as for $G^*(n, \mathbf{d})$. Hence $\tilde{G}^*(n, \mathbf{d}, \gamma)$ is simple if and only if $G^*(n, \mathbf{d})$ is. So, conditioned on $\tilde{G}^*(n, \mathbf{d}, \gamma)$ being simple, it is distributed as $\tilde{G}(n, \mathbf{d}, \gamma)$, and equation (21) implies that

$$\liminf \mathbb{P}(\tilde{G}^*(n, \mathbf{d}, \gamma) \text{ is simple}) > 0.$$

Therefore, we can prove Theorems 8 and 11 for one of the models $\tilde{G}^*(n, \mathbf{d}, \gamma)$ or $\tilde{G}(n, \mathbf{d}, \gamma)$, and it will imply that they are true for both.

5.2 Link between the graph $\tilde{G}(n, \mathbf{d}, \gamma)$ and the original graph $G(n, \mathbf{d})$

Let \tilde{G} be distributed as $\tilde{G}(n, \mathbf{d}, \gamma)$. We say that a vertex in $\tilde{G}(n, \mathbf{d}, \gamma)$ has parent $i \in \{1, \dots, n\}$ if it belongs to a clique that replaces the vertex i of $G(n, \mathbf{d})$ (when $X(i) = 1$) or if it is i (when $X(i) = 0$). For any subgraph $\tilde{H} \subset \tilde{G}$, we obtain the graph $\phi(\tilde{H})$ by identifying in \tilde{H} the vertices that have the same parent and that are connected in \tilde{H} . For instance, Figure 9 represents a clique of size 4 in \tilde{G} that comes from the replacement of a vertex i : thus all the vertices in the clique have the same parent i . In the subgraph \tilde{H} , some of the edges of the clique are not present (those in dots): the clique is split into two connected components. In the corresponding graph $\phi(\tilde{H})$, we merge the vertices of the clique that are connected together.

We use the same definition of $\phi(\tilde{H})$ when \tilde{H} is a subgraph of a random graph distributed as $\tilde{G}^*(n, \mathbf{d}, \gamma)$.

For any graph G , set $\nu(G)$ for the number of vertices in G . The next lemma will be useful in several proofs.

Lemma 14. *Let \tilde{G} be distributed as $\tilde{G}(n, \mathbf{d}, \gamma)$. Let H be a subgraph of $\phi(\tilde{G})$ such that $\nu(H) = o_p(n)$. Let \tilde{H} be the maximal subgraph of \tilde{G} such that $\phi(\tilde{H}) = H$. Then we have: $\nu(\tilde{H}) = o_p(\tilde{n})$.*

Proof. We can bound $\nu(\tilde{H})/n$ using Cauchy-Schwarz inequality:

$$\begin{aligned} \nu(\tilde{H})/n &\leq \sum_r r \nu_r(H)/n \\ &\leq \sqrt{\sum_r \frac{r^2 \nu_r(H)}{n}} \sqrt{\sum_r \frac{\nu_r(H)}{n}} \\ &\leq \sqrt{\frac{\sum_i d_i^2}{n}} \sqrt{\frac{\nu(H)}{n}} \end{aligned}$$

Yet Condition 1-(iii) implies that $\sum_i d_i^2/n = O(1)$, and by hypothesis $\nu(H)/n \rightarrow_p 0$, so $\nu(\tilde{H})/n \rightarrow_p 0$, and $\nu(\tilde{H})/\tilde{n} \rightarrow_p 0$ due to Proposition 2. \square

5.3 Proof of Theorem 8

A heuristic (using a branching process approximation) is given in Appendix A. We first give the idea of the proof.

Let \tilde{G} be distributed as $\tilde{G}^*(n, \mathbf{d}, \gamma)$ and $\pi \in [0, 1]$. In the percolated graph \tilde{G}_π , the removal of some edges inside a clique can split the clique into several connected components. In order to study the percolated graph \tilde{G}_π (as described in Section 3), we proceed in three steps:

- Step 1.* We consider only the edges that are *inside* a clique, and we delete independently each of them with probability $1 - \pi$. The graph we obtain is called $\tilde{G}_\pi^{(1)}$. We are interested in the graph $G' = \phi(\tilde{G}_\pi^{(1)})$. If we condition G' on its number of vertices n' and its degree sequence \mathbf{d}' , we have that G' is distributed as $G^*(n', \mathbf{d}')$. The first step consists in computing the asymptotic distribution of the degree sequence \mathbf{d}' .
- Step 2.* Then we delete independently with probability $1 - \pi$ each edge of G' , and we apply results of [13] in order to study the components' sizes in the percolated graph G'_π .
- Step 3.* We deduce the components' sizes in \tilde{G}_π from the previous step, using the fact that $\phi(\tilde{G}_\pi)$ is distributed as G'_π .

In the following, when we consider the model $\tilde{G}^*(n, \mathbf{d}, \gamma)$, we take the multiplicity of edges into account when we compute the degree of a vertex. More precisely, we say that a vertex in $G^*(n, \mathbf{d})$ or $\tilde{G}^*(n, \mathbf{d}, \gamma)$ has 'degree' d if it has d (simple) half-edges. For instance, each loop of a given vertex has contribution 2 in its degree.

Step 1. For $d \geq 1$, let $V_d^{(n)}$ be the set of vertices i in $G(n, \mathbf{d})$ with degree d and such that $X(i) = 1$: i is replaced by a clique $K(i)$ of size d in \tilde{G} . Let $K(i, \pi)$ be the subgraph of $K(i)$ obtained after a bond percolation with parameter π . We consider the subgraph $\tilde{F}_d(\pi) \subset \tilde{G}_\pi$ that contains the percolated version of the cliques with initial size d :

$$\tilde{F}_d(\pi) = \bigcup_{i \in V_d^{(n)}} K(i, \pi)$$

The next lemma gives the limit, as $n \rightarrow \infty$, for the number $N^{(n)}(d, k, \pi)$ of connected components in $\tilde{F}_d(\pi)$ whose size is $k \leq d$.

Lemma 15. *For any $d \geq 1$ and $k \leq d$, we have that:*

$$N^{(n)}(d, k, \pi)/n \xrightarrow{p} \frac{d}{k} f(d, k, \pi) p_d \gamma_d,$$

where $f(d, k, \pi)$ is given by (6).

Proof. For each vertex i in $V_d^{(n)}$, we label the vertices of $K(i, \pi)$ from 1 to d . We look at all the vertices with label 1, and we let $M^{(n)}(d, k, \pi)$ be the number of such vertices whose connected component in $K(i, \pi)$ has size k . Using the Law of Large Numbers and the fact that $|V_d^{(n)}|/n \rightarrow_p p_d \gamma_d$, we have that $M^{(n)}(d, k, \pi)/n \rightarrow_p f(d, k, \pi) p_d \gamma_d$, where $f(d, k, \pi)$ is by definition the probability that the component of 1 contains k vertices. So the total number of vertices in $\tilde{F}_d(\pi)$ that belongs to a component of size k is: $df(d, k, \pi) p_d \gamma_d n + o_p(n)$ and, in order to have the number of such components, we have to divide by k , which proves the lemma. \square

Let $\tilde{G}_\pi^{(1)}$ be the graph obtained from $\tilde{G}^*(n, \mathbf{d}, \gamma)$ when we replace each vertex i such that $X(i) = 1$ by the percolated clique $K(i, \pi)$. For any $k \geq 0$, let n'_k be the number of vertices with 'degree' k in the projected graph $G' = \phi(\tilde{G}_\pi^{(1)})$. In order to compute n'_k , we have to consider the vertices i such that

$X(i) = 0$ (there are $n_k - |V_k^{(n)}|$ such ones, where n_k is the number of vertices with 'degree' k in $G^*(n, \mathbf{d})$, and the vertices that come from a clique of initial size d , for some $d \geq k$ (each such vertex corresponds to a component of size k in $\tilde{F}_d(\pi)$, so there are $N^{(n)}(d, k, \pi)$ such ones). This gives the following relation, for all $k \geq 0$:

$$n'_k = n_k - |V_k^{(n)}| + \sum_{d \geq k} N^{(n)}(d, k, \pi),$$

So Lemma 15 gives the following asymptotic distribution for the degree sequence \mathbf{d}' :

Lemma 16. *Let $n' := \sum_k n'_k$ be the total number of vertices in G' . Then the proportion of vertices with degree k in G' has the following limit, as $n \rightarrow \infty$:*

$$\frac{n'_k}{n'} \xrightarrow{p} p'_k := \frac{\varrho_k}{\sum_\ell \varrho_\ell}$$

where $\varrho_k := p_k(1 - \gamma_k) + \sum_{d \geq k} \frac{d}{k} f(d, k, \pi) p_d \gamma_d$

In addition, the uniform summability of kn'_k/n implies the uniform summability of kn'_k/n' , so that $\sum_k kn'_k/n' \rightarrow_p \sum_k kp'_k$.

Step 2. We apply Theorem 3.9 in [13] to the random graph G' . Indeed, we can assume without loss of generality that the previous convergences (Lemma 16) hold a.s., and not just in probability (as in [13]: using the Skorohod coupling theorem, see [17] for instance, or arguing by selecting suitable subsequences). Then there is a giant component in the percolated graph G'_π if and only if

$$\pi \sum_d d(d-1)p'_d > \sum_d dp'_d,$$

which is equivalent to the fact that $\pi \mathbb{E}[\mathcal{K}(D^* + 1, \pi, \gamma) - 1] > 1$.

Step 3. The proof of (ii) follows easily from the previous step, and Lemma 14.

We give the main lines of the proof of (i). Assume $\pi > \pi_c$, which corresponds to $\pi \sum_d d(d-1)p'_d > \sum_d dp'_d$. Let \mathcal{C}_1 be the largest connected component in $G'_\pi = \phi(\tilde{G}_\pi)$ and $\check{\mathcal{C}}_1$ be the connected component of \tilde{G}_π such that $\phi(\check{\mathcal{C}}_1) = \mathcal{C}_1$.

We first compute the limit of $\nu(\check{\mathcal{C}}_1)/\tilde{n}$ as $n \rightarrow \infty$. Let g be the generating function

$$g(x) := \sum_k p'_k x^k = \frac{1}{\varrho} \sum_k \varrho_k x^k,$$

and recall that its mean is called $\mu = \sum_k k \varrho_k / \varrho$. Results in [13] show that the number $\nu_r(\mathcal{C}_1)$ of vertices with degree r in \mathcal{C}_1 satisfies: $\nu_r(\mathcal{C}_1)/n' \rightarrow_p \sum_{\ell \geq r} b_{\ell r}(\sqrt{\pi}) p'_\ell (1 - \xi^r)$, where n' is the total number of vertices in the graph G' , and ξ is the unique $\xi \in (0, 1)$ such that

$$g'(1 - \pi^{1/2} + \pi^{1/2}\xi) = \mu(1 - \pi^{-1/2} + \pi^{-1/2}\xi). \quad (22)$$

Since $h(z) = (1 - \pi + \pi z) \cdot g'(1 - \pi + \pi z)$, we have that ξ is the solution of (22) if and only if $\zeta = 1 - \pi^{-1/2} + \pi^{-1/2}\xi$ is the solution of $\mu\zeta(1 - \pi + \pi\zeta) = h(\zeta)$. Note that we used the second notation in the statement of Theorem 8, in order to be consistent with the notations of Theorem 10 (in fact, we could have used either results in [13] or Theorem 11 of [20] for the current proof).

Unfortunately, we cannot deduce directly the size of $\check{\mathcal{C}}_1$ using only the asymptotic for $\nu_r(\mathcal{C}_1)/n'$, $r \geq 0$. We have to be more precise: let $\nu_r^0(\mathcal{C}_1)$ (resp. $\nu_r^1(\mathcal{C}_1)$) be the number of vertices i with degree r in \mathcal{C}_1 such that $X(i) = 0$ (resp. $X(i) = 1$). Then we have:

$$\begin{aligned} \nu_r^0(\mathcal{C}_1)/n' &\rightarrow_p \sum_{\ell \geq r} b_{\ell r}(\sqrt{\pi}) p_\ell (1 - \gamma_\ell) (1 - \xi^r) / \varrho, \\ \nu_r^1(\mathcal{C}_1)/n' &\rightarrow_p \sum_{\ell \geq r} b_{\ell r}(\sqrt{\pi}) \sum_{d \geq \ell} \frac{d}{\ell} f(d, \ell, \pi) p_d \gamma_d (1 - \xi^r) / \varrho. \end{aligned}$$

In summations, d represents the degree of vertices in the initial graph \tilde{G} , ℓ the degree of vertices in $\tilde{G}_\pi^{(1)}$ (after the percolation *inside* cliques), and r the degree of vertices after the percolation on *external* edges. In order to recover $\nu(\check{\mathcal{C}}_1)$, we have to multiply each term in $\nu_r^1(\mathcal{C}_1)$ by ℓ , and then sum over all r , which gives (exchanging summations on r and ℓ):

$$\nu(\check{\mathcal{C}}_1)/n' \xrightarrow{p} \frac{1}{\varrho} \sum_{k \geq 1} \sigma_k \left(1 - (1 - \pi^{1/2} + \pi^{1/2} \xi)^k \right).$$

Using that $n'/n \rightarrow_p \varrho$ and $\tilde{n}/n \rightarrow_p \tilde{\gamma}$, we obtain that $\nu(\check{\mathcal{C}}_1)/\tilde{n} \rightarrow_p L(\zeta)$.

Let $\tilde{\mathcal{C}}_1$ be the largest component in \tilde{G}_π . Adding cliques changes the sizes of the connected components: hence we have to prove that $\check{\mathcal{C}}_1 = \tilde{\mathcal{C}}_1$ whp. Let $\tilde{\mathcal{C}}$ be any other component of \tilde{H} different from $\tilde{\mathcal{C}}_1$. Its projection $\mathcal{C} = \phi(\tilde{\mathcal{C}})$ is different from \mathcal{C}_1 , so Theorem 3.9 in [13] implies that $\nu(\mathcal{C})/n \rightarrow_p 0$. Using Lemma 14 with $H = \mathcal{C}$ shows that $\nu(\tilde{\mathcal{C}})/\tilde{n} \rightarrow_p 0$. Hence $\check{\mathcal{C}}_1$ is the largest connected component of \tilde{G} whp, which ends the proof.

5.4 Proof of Theorem 10

The difference with the previous proof is the following: instead of using Theorem 3.9 of [13] in steps 2 and 3, we use Theorem 10 of [20].

Indeed the first step is the same: the graph $G' = \phi(\tilde{G}_\pi^{(1)})$ (where $\tilde{G}_\pi^{(1)}$ is the graph obtained from $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$ after a bond percolation on the edges *inside* cliques only) has asymptotic degree distribution $\mathbf{p}' = (p'_k)_k$, with $p'_k = \varrho_k/\varrho$.

We apply (a slight extension of) Theorem 10 in [20] for the graph G' (with $t_{s\ell} = \mathbf{1}_{\{\ell=0\}}$). Let ν_s^0 be the number of vertices i such that $X(i) = 0$ and that satisfy: the degree of i in G' (that is to say before the bond percolation in G') is s , and i is active at the end of the process. Let ν_{ds}^1 be the number of vertices i such that $X(i) = 1$ and that satisfy: the degree of i in the original graph $\phi(\tilde{G})$ is d , the degree of i in G' is $s \leq d$, and i is active at the end of the process. The probability that such a node i (with degree d in $\phi(\tilde{G})$ and s in G') does not belong to the original seed S is $(1 - \alpha_d)^s$ (and initial activations are independent among nodes). Hence we have:

$$\begin{aligned} \nu_s^0/n' &\rightarrow_p p_s(1 - \gamma_s) [1 - (1 - \alpha_s)(1 - \pi + \pi\zeta)^s] / \varrho, \\ \nu_{ds}^1/n' &\rightarrow_p \frac{d}{s} f(d, s, \pi) \gamma_d p_d [1 - (1 - \alpha_d)^s (1 - \pi + \pi\zeta)^s] / \varrho, \end{aligned}$$

where ζ is given by (12). In order to obtain $C^b(\pi, \boldsymbol{\alpha})$, we have to multiply ν_{ds}^1 by s , and sum over all d and s , which gives:

$$\begin{aligned} C^b(\pi, \boldsymbol{\alpha})/n' &\rightarrow_p \sum_s p_s(1 - \gamma_s) [1 - (1 - \alpha_s)(1 - \pi + \pi\zeta)^s] / \varrho \\ &\quad + \sum_{d \geq s} d f(d, s, \pi) \gamma_d p_d [1 - (1 - \alpha_d)^s (1 - \pi + \pi\zeta)^s] / \varrho \end{aligned}$$

and ends the proof (since $n'/n \rightarrow_p \varrho$ and $\tilde{n}/n \rightarrow_p \tilde{\gamma}$).

5.5 Proof of Theorem 11

As before, we say that a vertex in $\tilde{G}(n, \mathbf{d}, \boldsymbol{\gamma})$ has parent $i \in \{1, \dots, n\}$ if it belongs to a clique that replaces the vertex i of $G(n, \mathbf{d})$ (when $X(i) = 1$) or if it is i (when $X(i) = 0$). For any graph G and any vertex v of G , let $D(v, \mathbf{t})$ be the subgraph of G induced by the final set of active vertices, when v is the only vertex in the initial seed. With the notations of 4.2, the number of vertices in $D(v, \mathbf{t})$ is $C(v, \mathbf{t})$. When

H is a subgraph of G , we set $D(H, \mathbf{t})$ for the subgraph induced by the final set of active vertices in G , when the initial vertices in the seed are those of H .

We can make a comparison between an epidemic starting from a vertex u in $\tilde{G}(n, \mathbf{d}, \gamma)$, and the epidemic that would have been generated by its parent i in $G(n, \mathbf{d})$ (recall that thresholds are drawn such that a vertex u has the same threshold as its parent i).

Proposition 17. *Let u be a vertex of $\tilde{G} = \tilde{G}(n, \mathbf{d}, \gamma)$. Let i be its parent, and K be the clique generated by i if $X(i) = 1$ (otherwise, set $K = \{u\}$). Then we can bound the epidemic generated by u the following way:*

$$\phi(D(u, \mathbf{t})) \subset \phi(D(K, \mathbf{t})) \subset D(i, \mathbf{t}).$$

The proof follows from Lemma 12 in 4.2. In addition, we have the following lemma, that is a consequence of [20]:

Lemma 18. *Assume $\sum_r r(r-1)p_r t_{r0} < \sum_r r p_r$. Let u be a vertex chosen uniformly at random among the vertices of $\tilde{G} = \tilde{G}(n, \mathbf{d}, \gamma)$, and let i be the parent of u . Then the size of the epidemic generated by i in $\phi(\tilde{G})$ is $C(i, \mathbf{t}) = o_p(n)$.*

Proof. We cannot use directly the result of [20] that says that, if $\sum_r r(r-1)p_r t_{r0} < \sum_r r p_r$ and i is chosen uniformly at random in $G(n, \mathbf{d})$, then $C(i, \mathbf{t}) = o_p(n)$. The idea is to apply Theorem 10 [20], with a parameter $\alpha = (\alpha_d)_{d=0}^\infty$ that satisfies: $\alpha_d = (d\gamma_d + 1 - \gamma_d)\alpha$ for all d , α being a positive constant. Then the same arguments as for the proof of Theorem 11-(ii) [20] work. \square

This allows to deduce easily the case (ii) of Theorem 11: assume $\sum_r r(r-1)p_r t_{r0} < \sum_r r p_r$. Then combining Proposition 17 and Lemma 18 gives that the number of vertices in $\phi(D(u, \mathbf{t}))$ is $o_p(n)$ if u is chosen uniformly at random in \tilde{G} . Applying Lemma 14 with $H = D(u, \mathbf{t})$ concludes the proof of case (ii).

We now assume that the cascade condition (14) is satisfied.

The proof of (16) is a consequence of a result from [20]. Indeed let \tilde{G} be distributed as $\tilde{G}(n, \mathbf{d}, \gamma)$. Let \tilde{H} (resp. H) be the subgraph of \tilde{G} (resp. $\phi(\tilde{G})$) induced by the vertices of threshold zero. Note that $\phi(\tilde{H}) = H$. We use Theorem 11 in [20] for the graph $\phi(\tilde{G})$ (with parameter $\pi = 1$): it gives the components' sizes in H . Let \mathcal{C}_1 (resp. $\tilde{\mathcal{C}}_1$) be the largest connected component in H (resp. \tilde{H}). The number $\nu_r(\mathcal{C}_1)$ of vertices with degree r in \mathcal{C}_1 is computed in the proof of Theorem 11 in [20]: $\nu_r(\mathcal{C}_1)/n \rightarrow_p p_r t_{r0}(1 - \xi^r)$, where ξ is defined in (15). Hence we can deduce the size of the connected component $\tilde{\mathcal{C}}_1$ in \tilde{H} such that $\phi(\tilde{\mathcal{C}}_1) = \mathcal{C}_1$: $\nu(\tilde{\mathcal{C}}_1)/\tilde{n} \rightarrow_p \sum_d [d\gamma_d + (1 - \gamma_d)] p_d t_{d0}(1 - \xi^d)/\tilde{\gamma}$. The way to show that $\tilde{\mathcal{C}}_1 = \tilde{\mathcal{C}}_1$ whp is similar to the end of Theorem 8. This ends the proof of (16).

The idea for the rest of the proof is to make a coupling between the epidemic on \tilde{G} (with threshold distribution \mathbf{t}), and an epidemic on $\phi(\tilde{G})$ with a different threshold distribution, that we call $\mathbf{t}' = (t'_{s\ell})_{s,\ell}$.

Proposition 19. *Assume the epidemic on $\tilde{G} = \tilde{G}(n, \mathbf{d}, \gamma)$ starts from a vertex u that has threshold zero, and let i be the parent of u in $\phi(\tilde{G})$. We consider the following distribution of thresholds $(t'_{s\ell})_{0 \leq \ell \leq s}$ for each $s \geq 0$:*

- $t'_{s0} = t_{s0}$;
- $t'_{s\ell} = (1 - \gamma_s)t_{s\ell}$ for all $0 < \ell < s$;
- $t'_{ss} = (1 - \gamma_s)t_{ss} + \gamma_s(1 - t_{s0})$.

Then there exist random thresholds $(k'(j))_{1 \leq j \leq n}$ with the distribution $\mathbf{t}' = (t'_{s\ell})_{0 \leq \ell \leq s}$ defined above such that

$$\phi(D(u, \mathbf{t})) = D(i, \mathbf{t}'),$$

where $D(i, \mathbf{t}')$ is the subgraph induced by the final set of active vertices in the symmetric threshold model starting from i in $\phi(\tilde{G})$, with threshold distribution $\mathbf{t}' = (t'_{s\ell})_{s,\ell}$.

Proof. Note that each vertex i of $\phi(\tilde{G})$ has two thresholds: $k(i)$ (that we used to define the epidemic on \tilde{G}), and the new threshold $k'(i)$, that we will define to use to make a comparison with the epidemic on \tilde{G} . Until the end of the proof, if no precision is given, we refer to $k'(i)$ as the threshold of i .

We explicit the natural coupling between the symmetric threshold model with parameter \mathbf{t} in \tilde{G} and the one with parameter \mathbf{t}' in $\phi(\tilde{G})$. If u belongs to a clique, then the whole clique becomes active at the next step, so we can start the epidemic in $\phi(\tilde{G}) = G(n, \mathbf{d})$ from the parent i of u . Let v be the neighbor of u outside the clique, and let j be the parent of v . If the threshold $k(v)$ of vertex v in \tilde{G} is zero, then v (and its whole clique if it has one) becomes active (see Lemma 12). In this case, we choose $k'(j) := 0$ for the threshold of j in $\phi(\tilde{G})$, so that it becomes also active in $\phi(\tilde{G})$. If $k(v) > 0$, then there are two cases:

- If $X(j) = 1$, vertex v and its clique stay inactive (Lemma 12). In this case, we choose $k'(j) := s$ for the threshold of j (so that it stays inactive).
- If $X(j) = 0$, vertex v becomes active if and only if it has at least $k(v) + 1$ active neighbors. So we set $k'(j) := k(v) = k(j)$ for the threshold of j .

Since the random variables $X(j)$, for j in $\phi(\tilde{G})$, are independent, the thresholds we associate to each node are also independent. In addition, we can easily verify that the conditional probability distribution of thresholds (knowing that the degree of the node is s) is given by $(t'_{s\ell})_{0 \leq \ell \leq s}$. In fact, the epidemic we consider in $\phi(\tilde{G})$ is almost the same as the one with parameter \mathbf{t} , except that we randomly put some nodes j (those such that $X(j) = 1$) to a threshold so high that they stay inactive. \square

More precisely, let $C_{s\ell}(u, \mathbf{t})$ (resp. $C'_{s\ell}(i, \mathbf{t}')$) be the final number of active vertices with degree $s \geq 0$ and threshold ℓ at the end of the symmetric threshold epidemic on \tilde{G} (resp. $\phi(\tilde{G})$), with threshold parameter \mathbf{t} (resp. \mathbf{t}'), when the only vertex in the initial seed is u (resp. i). Then, using the coupling described above, we have the following result, for each degree $s \geq 0$:

- $C_{s0}(u, \mathbf{t}) = C'_{s0}(i, \mathbf{t}')[sY_s + (1 - Y_s)]$, where Y_s is the proportion of vertices j in $\phi(\tilde{G})$ such that $X(j) = 1$, among those that have degree s and that belong to the cascade triggered by i .
- For all $\ell \neq 0$, we have that $C_{s\ell}(u, \mathbf{t}) = C'_{s\ell}(i, \mathbf{t}')$, since the vertices of positive threshold that belong to the cascade triggered by u are exactly those that are not replaced by a clique.

We have that $Y_s/n \rightarrow_p \gamma_s$ for all s , and the limit for $C'_{s\ell}(i, \mathbf{t}')$ is given by the following lemma (which is a slight extension of Theorem 11 [20]):

Lemma 20. *Assume (using the notations of Theorem 11) that $\zeta = 0$ or ζ is such that there exists $\varepsilon > 0$ with $\lambda z^2 < h(z)$ for $z \in (\zeta - \varepsilon, \zeta)$. Then, for any i that belongs to the set of pivotal players in $\phi(\tilde{G})$, we have:*

$$C'_{s\ell}(i, \mathbf{t}')/n \rightarrow_p p_s t'_{s\ell} \left(1 - \sum_{r \geq s-\ell} b_{sr}(\zeta) \right).$$

In particular, for $\ell = 0$, we have: $C'_{s0}(i, \mathbf{t}')/n \rightarrow_p p_s t'_{s0} (1 - \zeta^s)$.

Proof. By slight extension of Theorem 11 [20], the number of inactive nodes with original degree s , degree r in the graph of inactive nodes and threshold ℓ tends to $\sum_{i \geq s-r-\ell} p_s t'_{s\ell} b_{sr}(\zeta) b_{s-r,i}(0) = p_s t'_{s\ell} b_{sr}(\zeta) \mathbf{1}\{r \geq s-\ell\}$. Hence summing over r gives that the number of inactive nodes with original degree s and threshold ℓ tends to $p_s t'_{s\ell} \sum_{r \geq s-\ell} b_{sr}(\zeta)$, which ends the proof. \square

We assume that $\zeta = 0$ or ζ is such that there exists $\varepsilon > 0$ with $\lambda z^2 < h(z)$ for $z \in (\zeta - \varepsilon, \zeta)$. Let u be a vertex in \tilde{G} whose parent i belongs to the set of pivotal players in $\phi(\tilde{G})$. Let $C_s(u, \mathbf{t})$ be the final number of active vertices with degree $s \geq 0$ at the end of the symmetric threshold epidemic on \tilde{G} , with threshold parameter \mathbf{t} , when the only vertex in the initial seed is u . Then we have:

$$\frac{C_s(u, \mathbf{t})}{\tilde{n}} \xrightarrow{p} \frac{[s\gamma_s + (1 - \gamma_s)]p_s}{\tilde{\gamma}} t'_{s0}(1 - \zeta^s) + \frac{p_s}{\tilde{\gamma}} \sum_{\ell \neq 0} t'_{s\ell} \left(1 - \sum_{r \geq s-\ell} b_{sr}(\zeta) \right)$$

Using the definition of \mathbf{t}' , we have that $t'_{s0} = t_{s0}$ and:

$$\sum_{\ell \neq 0} t'_{s\ell} \sum_{r \geq s-\ell} b_{sr}(\zeta) = \sum_{\ell \neq 0} (1 - \gamma_s) t_{s\ell} \sum_{r \geq s-\ell} b_{sr}(\zeta) + \gamma_s (1 - t_{s0}),$$

which finally gives:

$$\frac{C_s(u, \mathbf{t})}{\tilde{n}} \xrightarrow{p} \frac{[s\gamma_s + (1 - \gamma_s)]p_s}{\tilde{\gamma}} t_{s0}(1 - \zeta^s) + \frac{(1 - \gamma_s)p_s}{\tilde{\gamma}} \left(1 - t_{s0} - \sum_{\ell \neq 0} t_{s\ell} \sum_{r \geq s-\ell} b_{sr}(\zeta) \right)$$

Then, by an argument similar as the one at the end of Theorem 8 or equation (16), we have that u belongs to the set of pivotal players in \tilde{G} , which ends the proof.

5.6 Proof of Theorem 13

We use the same idea as in the previous proof. The same statement as for Proposition 19 holds when the epidemic starts from a set (instead of a single vertex u). Indeed, let S be the initial seed in $\phi(\tilde{G})$. By definition, the initial seed \tilde{S} in \tilde{G} consists of the vertices whose parent belongs to S .

Let $C_{s\ell}(\tilde{S}, \mathbf{t})$ (resp. $C'_{s\ell}(S, \mathbf{t}')$) be the final number of active vertices with degree $s \geq 0$ and threshold ℓ at the end of the symmetric threshold epidemic on \tilde{G} (resp. $\phi(\tilde{G})$), with threshold parameter \mathbf{t} (resp. \mathbf{t}' , defined in Proposition 19), when the initial seed is \tilde{S} (resp. S).

Using a slight extension of Theorem 10 in [20], we have, for all $s \geq 0$ and $\ell \geq 0$:

$$C'_{s\ell}(S, \mathbf{t}')/n \rightarrow_p p_s t'_{s\ell} \left(\alpha_s + (1 - \alpha_s) \left(1 - \sum_{r \geq s-\ell} b_{sr}(\zeta) \right) \right),$$

where ζ is defined in (20). More precisely, the first term $p_s t'_{s\ell} \alpha_s$ comes from the vertices that belong to the initial seed S , and the second one $p_s t'_{s\ell} (1 - \alpha_s) \left(1 - \sum_{r \geq s-\ell} b_{sr}(\zeta) \right)$ comes from those that are activated during the process. In order to obtain the asymptotic for $C_{s\ell}(\tilde{S}, \mathbf{t})/n$, we have to multiply the first term by $(s\gamma_s + 1 - \gamma_s)$. The multiplicative constant for the second term depends on the value of the threshold ℓ : if $\ell = 0$, we multiply the second term by $(s\gamma_s + 1 - \gamma_s)$, and if $\ell > 0$, we multiply it by 1 (since the vertices with positive threshold that are activated during the process necessarily do not belong to a clique). Summing over s, ℓ and replacing \mathbf{t}' by its expression gives the following limit, as $n \rightarrow \infty$:

$$\begin{aligned} C(\mathbf{t}, \boldsymbol{\alpha})/n &\rightarrow_p \sum_s p_s t_{s0} (s\gamma_s + 1 - \gamma_s) \left[\alpha_s + (1 - \alpha_s) (1 - \zeta^s) \right] \\ &\quad + \sum_s p_s (1 - \gamma_s) \alpha_s (s\gamma_s + 1 - \gamma_s) (1 - t_{s0}) \\ &\quad + \sum_s p_s (1 - \gamma_s) (1 - \alpha_s) \left[(1 - t_{s0}) - \sum_{\ell \neq 0} t_{s\ell} \sum_{r \geq s-\ell} b_{sr}(\zeta) \right] \\ &\quad + \sum_s p_s \gamma_s (1 - t_{s0}) \alpha_s (\gamma_s + 1 - \gamma_s). \end{aligned}$$

Gathering some terms and using that $\tilde{n}/n \rightarrow_p \tilde{\gamma}$ ends the proof of Theorem 13.

6 Conclusions

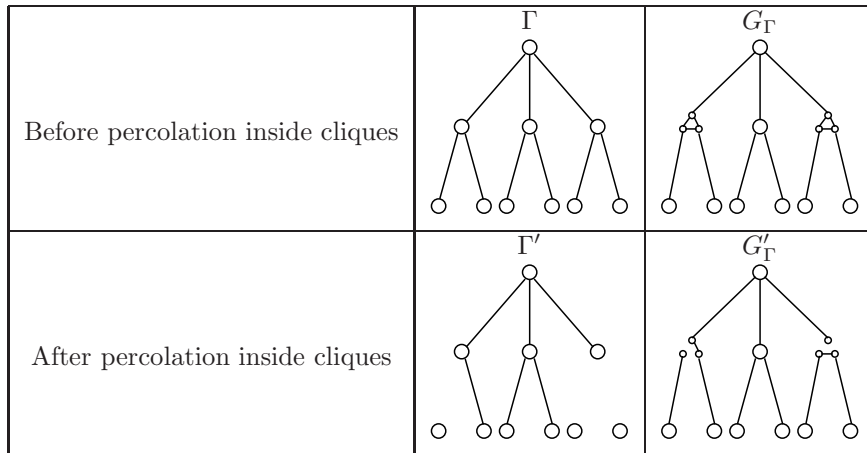
Up to our knowledge, our analysis is the first systematic study of random graphs with both a tunable asymptotic degree distribution and a clustering coefficient. Our model has the advantage to still be tractable for the analysis of diffusion or symmetric threshold model.

For both models, we are able to derive explicit formulas for the cascade condition, *i.e.* the condition under which a single infected individual can turn a positive fraction of the population into infected individuals. When such a cascade is possible, the expression of its size is given analytically. In the case of random regular graphs, we proved that the clustering ‘inhibits’ the diffusion process. Numerical evaluations also show that clustering decreases the cascade size of the diffusion process for regular graphs, and ‘inhibits’ the diffusion process for power-law graphs. The impact of clustering on the symmetric threshold model is studied in the particular case of the contagion model described in 4.3: numerical evaluations show that the effect of clustering on the contagion process depends on the value of the mean degree in the graph: while clustering ‘inhibits’ the contagion for a low mean degree, the contrary happens in the high values regime. When a cascade is possible, we observe that clustering decreases its size.

In addition, we can also compute explicitly the cascade size in the case of a degree based activation, for both diffusion and symmetric threshold models. This theoretical analysis paves the way to a possible control of such epidemic processes as done in [6] or [19].

A Branching process approximation for the diffusion threshold (with a single activation)

We can guess the value of the diffusion threshold π_c given in Theorem 8 using a branching process approximation. Indeed the random graph $G(n, \mathbf{d})$ can be approximated by a branching process Γ in which each node (except the root) has a number of offspring distributed as D^* . We add cliques in this branching process as in $\tilde{G}(n, \mathbf{d}, \gamma)$, which gives a graph G_Γ . We then proceed in two steps: first we delete independently with probability $1 - \pi$ (in G_Γ) each ‘internal’ edge, *i.e.* edge inside a clique; second we delete independently with probability $1 - \pi$ (in the new graph) each ‘external’ edge, *i.e.* edge outside cliques.



After the first deletion of edges, we get a new graph G'_Γ in which original cliques can be broken into several components. If we make the equivalent transformation in the original branching process Γ , it

means that a node can lose some of its children.

More precisely, we consider a node i in the branching process Γ . Let e be the edge of i that links i to the previous generation. The degree of i is distributed as $D^* + 1$. We assume $D^* + 1 = d$. In G_Γ , node i is replaced by a clique K with probability γ_d . In that case, let v be the vertex in K whose edge outside the clique is e . After having deleted independently each edge inside the clique with probability $1 - \pi$, the probability that the component of vertex v inside K contains k vertices (including v itself) is given by $f(d, k, \pi)$. Hence the probability that v is linked to k vertices (including the one linked by e) is: $(1 - \gamma_d)\mathbb{1}(d = k) + \gamma_d f(d, k, \pi) = \mathbb{P}(\mathcal{K}(D^* + 1, \pi, \gamma) = k)$. The new distribution of offspring in the corresponding branching process Γ' is thus $\mathcal{K}(D^* + 1, \pi, \gamma) - 1$. Finally, we remove each (external) edge with probability $1 - \pi$, which gives $\pi \mathbb{E}[\mathcal{K}(D^* + 1, \pi, \gamma) - 1]$ for the expected number of offspring.

References

- [1] D. Acemoglu, A. Ozdaglar, and E. Yildiz. Diffusion of innovations in social networks. In *IEEE Conference on Decision and Control (CDC)*, 2011.
- [2] H. Amini. Bootstrap percolation and diffusion in random graphs with given vertex degrees. *The Electronic Journal of Combinatorics*, 17, 2010.
- [3] J. Balogh and B. G. Pittel. Bootstrap percolation on the random regular graph. *Random Structures Algorithms*, 30(1-2):257–286, 2007.
- [4] L. E. Blume. The statistical mechanics of best-response strategy revision. *Games Econom. Behav.*, 11(2):111–145, 1995. Evolutionary game theory in biology and economics.
- [5] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.
- [6] C. Borgs, J. Chayes, A. Ganesh, and A. Saberi. How to distribute antidote to control epidemics. *Random Struct. Algorithms*, 37:204–222, September 2010.
- [7] T. Britton, M. Deijfen, A. N. Lagerås, and M. Lindholm. Epidemics on random graphs with tunable clustering. *J. Appl. Probab.*, 45(3):743–756, 2008.
- [8] E. Coupechoux and M. Lelarge. Impact of clustering on diffusions and contagions in random networks. In *NetGCoop 2011: International conference on NETWORK Games, COntrol and OPTimization*, 2011.
- [9] M. Deijfen and W. Kets. Random intersection graphs with tunable degree distribution and clustering. *Probab. Eng. Inform. Sci.*, 23:661–674, 2008.
- [10] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30:1141–1144, 1959.
- [11] J. P. Gleeson, S. Melnik, and A. Hackett. How clustering affects the bond percolation threshold in complex networks. *Physical Review E*, 81, 2010.
- [12] M. Granovetter. Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.
- [13] S. Janson. On percolation in random graphs with given vertex degrees. *Electron. J. Probab.*, 14:no. 5, 87–118, 2009.
- [14] S. Janson. The probability that a random multigraph is simple. *Combinatorics, Probability and Computing*, 18(1-2):205–225, 2009.
- [15] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

- [16] M. Kaiser. Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. *New Journal of Physics*, 10, 2008.
- [17] O. Kallenberg. *Foundations of modern probability*. Springer-Verlag, 2002.
- [18] J. Kleinberg. *Cascading Behavior in Networks: Algorithmic and Economic Issues*. Cambridge University Press, 2007.
- [19] M. Lelarge. Efficient control of epidemics over random networks. In J. R. Douceur, A. G. Greenberg, T. Bonald, and J. Nieh, editors, *SIGMETRICS/Performance*, pages 1–12. ACM, 2009.
- [20] M. Lelarge. Diffusion and cascading behavior in random networks. *under revision for Games and Economic Behavior*, arxiv/1012.2062, 2010.
- [21] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, 6(2-3):161–179, 1995.
- [22] S. Morris. Contagion. *Rev. Econom. Stud.*, 67(1):57–78, 2000.
- [23] M. E. J. Newman. Properties of highly clustered networks. *Phys. Rev. E*, 68(2):026121, Aug 2003.
- [24] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.*, 45(2):167–256 (electronic), 2003.
- [25] M. E. J. Newman. Random graphs with clustering. *Phys. Rev. Lett.*, 2009.
- [26] P. Trapman. On analytical approaches to epidemics on networks. *Theoretical Population Biology*, 71(2):160–173, 2007.
- [27] F. Vega-Redondo. *Complex social networks*, volume 44 of *Econometric Society Monographs*. Cambridge University Press, Cambridge, 2007.
- [28] D. J. Watts. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. USA*, 99(9):5766–5771 (electronic), 2002.
- [29] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.